Modelling health related behaviours using geodemographics: applications in social marketing and preventative health

Catherine Emma Jones,

Department of Geography,

University College London.

May 2008

This thesis is submitted in partial fulfilment for the requirements of the degree Doctor of Philosophy (PhD)

Author's Declaration

I, Catherine Emma Jones confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The increased incidence of lifestyle related diseases, such as obesity and diabetes, across the western world is now an established fact, and presents many challenges to researchers trying to understand the determinants of poor health. Measurement of health needs and health outcomes is a fundamental component of evidence-based policy, strategy and delivery of health care services and interventions at scales from the local to the national. A central contention of this thesis is that health outcome indicators should be cognisant of factors such as personal behaviour, lifestyles, community influences, living and working conditions, accessibility to services and educational attainment which all impact upon the health of the individual and the wider community. It is therefore sensible to explore these differences by understanding both the social space comprising of different population sub-groups and the geographical within which they live.

Good quality data underlie the functioning of evidence-based decisions. Data provide the building blocks for understanding the nature and composition of neighbourhoods, together with the expected health outcomes of their residents. But within the health arena there are many complicated data issues. Existing operational health data sets are often incomplete or not up-to-date and accessibility is often limited by data protection and medical confidentiality policies. They are derived from disparate sources: GP registers, Hospital Episode statistics (people who are admitted to hospital), Child Registry and Accident and Emergency records, all adhering to different data collection and storage standards and systems that vary between organisations. Cross-referencing between these datasets is technically difficult because of these issues. Frequent quality issues of operational health data limit the extent of analysis that can be carried out with confidence.

Furthermore, health survey data are released at coarse geographical scales where the ecological fallacy limits the potential for exploring local variability.

Given these limiting factors, the theme of this research is to extend the health inequalities research and its associated data framework to explore variability in the spatial and social domain This enables the identification of social facts relating to health harming lifestyle choices and behaviours that contribute to 'diseases of comfort'. This is carried out by developing and exploring the usefulness of geodemographics for analysing health inequalities, thereby adding the social and spatial context to our understanding of causes of health inequalities.

This thesis presents a more straightforward yet effective alternative to exploring the measurement of health impacting behaviours and predicting health outcomes using operational health data, national health surveys and a geodemographic classification. Geodemographic analysis of health outcomes can capture different lifestyle behaviours, and has already proven useful not only in improving customer segmentation in the commercial sector, but also to better target public services (Harris et al., 2005). By applying geodemographic classifications to national health surveys and NHS operational datasets at postcode level, interesting conclusions can be drawn in terms of different health harming lifestyle behaviours at very fine scales.

Furthermore it is common practice that academic research projects occur in isolation, and exploitation of research findings and best practices in local government sectors is often beset by many obstacles. Consequently, within local government the adoption of new innovative techniques and tools may often be slow. An inner London Primary Care Trust (PCT) is used as a test bed for disseminating and evaluating the geodemographic framework and indicators. The concluding sections of the thesis discuss the practicalities of embedding geodemographics in particular and geography in general into a professional environment where these technologies are new and innovative.

Acknowledgements

There are a number of people whose help, support and guidance I feel indebted to, and without which I surely would not have been able to finish. I must first extend my greatest thanks and appreciation to my supervisors Prof. Paul Longley and Dr. Muki Haklay. Their support and infinite wisdom has provided me with guidance all along this journey. I am privileged to have been their student.

Thanks must be expressed to my co-funders of Camden Primary Care Trust (PCT) and the Economic and Social Research Council (ESRC) for funding for this research through the Knowledge Transfer Partnership, KTP-037. My colleagues in the Health Intelligence Team provided witty conversation throughout lunch times and gave much encouragement. Particular thanks should be expressed to Kawser Zannath and Pablo Mateos with whom I shared an office. Richard Webber and Emily Sparks of Experian must be thanked for providing me with a rich data source. My PhD companions at UCL have shared wonderful bits of advice and experience in particular Claire Ellul.

My final thanks go to my friends and family who have endured the three years or so of the thesis writing with me. Their unconditional support and friendship has both supported and encouraged me during the highs and the lows. John, you have been wonderful, thank you so much for everything. Mum, Dad & Mark - Love you! Diolch yn fawr. Andreé and Jean Claude - villmols merci! Patrick your kindness, patience, support and love I cherish dearly. I promise to have time to learn Luxembourgish now!!

List of Abbreviations and Acronyms

ACORN	A Classification of Residential Neighbourhoods
AFPD	All Fields Postcode Directory
AGI	Association of Geographic Information
BMI	Body Mass Index
CBD	Central Business District
CELBSS	Central and East London Breast Screening Services
CHD	Coronary Heart Disease
COPD	Chronic Obstructive Pulmonary Disease
DoH	Department of Health
GIS	Geographical Information Systems/Science
GPs	General Practitioner
GWR	Geographically Weighted Regression
HES	Hospital Episode Statistics
HM	Her Majesty
HSE	Health Survey for England
IACC	Intra Area Correlation Coefficient
ICD	International Classification of Diseases
IVF	In vetro fertilisation
IMD	Index of Multiple Deprivation
KDE	Kernel Density Estimation
KTP	Knowledge Transfer Partnership
LISA	Local Indicators of Spatial Association
MAUP	Modifiable Areal Unit Problem
NATCEN	National Centre of Social Research
NCC	National Consumer Council
NCLSHA	North Central London Strategic Health Authority
NHS	National Health Service
NHSBSP	NHS Breast Screening Programme
NPfIT	National Programme for IT
NS	Non Smokers
NSMC	National Social Marketing Centre
OA	Output Area
OAC	Output Area Classification
OED	Oxford English Dictionary
ONS	Office of National Statistics
PBS	Public Health Observatory, Brent PCT and the School of Health and Related

	Research
PC	Postcode
PCT	Primary Care Trust
QOF	Quality or Outcomes Framework
ScHARR	School of Health and Related Research
SHA	Strategic Health Authorities
SM	Smokers
SOA	Super Output Area (SOA)
UCL	University College London
UK	United Kingdom
YHPHO	Yorkshire & Humber Public Health Observatory
YPLL	Years of potential life loss

Glossary of Terms

Term	Description
Age cohort	A group of people within the same age band.
Bonding capital	Bonding capital refers to ties between people in similar situations within a close network.
Bridging capital	Bridging capital encompasses distant ties of like persons, such as loose friendships and workmates.
Case mix	Different types of cases that results from different types of patients.
Causal path	The major factors, determinants and links that led to disease and morbidity.
Commissioning	Health commissioning is the means by which best value for patients and tax payers is secured, ensuring best possible health outcomes and best possible healthcare,.
Coverage rate	A coverage rate is the quantitative measure that defines the proportion of eligible women who have received screening at least once in the previous three years.
Deprivation measures	Composite aggregate indicators which provide comparable measures of deprivation for small areas.
Disease aetiology	Disease aetiology considers the causes, origins, evolution, and implications of different diseases.
Diseases of comfort	Diseases of comfort are diseases which have emerged as the result of living in a modern society and correspond to chronic (long-term) illnesses associated with obesity, lack of physical activity and lifestyle choices.
Ecological fallacy	The issues that arise from making inferences about the individual based upon patterns observed at the group level.
External competition	External competition relates to the factors competing for time and attention of an audience.
Formal scale	Formal scale corresponds to the organisational scale of the NHS or aspatial administrative units.
Functional scale	Functional scale represents the social scale of the neighbourhood.
Geodemographic / neighbourhood Group	The Mosaic (Experian: Nottingham) commercial geodemographic classification segments the UK postcodes into 61 different unique Types of population. These population Types are then aggregated into hierarchical Groups, of which there are 11.
Geodemographic / neighbourhood Type	The Mosaic (Experian: Nottingham) commercial geodemographic classification segments the UK postcodes into 61 different unique Types of population that represent shared socio-economic and lifestyle characteristics of populations within each Type.
Geodemographic /neighbourhood typology	Systematic organisation of different sections of the population into types according to shared characteristics.
Geodemographic classification	Small area measure of shared socio-economic and lifestyle conditions based on social similarity and proximity.
Geographically weighted regression	Geographically weighted regression is an exploratory regression technique that takes into account the spatial data and location.

Term	Description
Gradients of health inequalities	Health Inequalities do not just exist between the most affluent and the most deprived, inequalities exist as a gradient traversing the social spectrum.
Health behaviours	Individual behaviours that influence ones health outcomes. Positive health behaviours protect, modify and promote health. Negative health behaviours do the inverse.
Health inequalities	Differences in health status or social determinants that result in gaps between different population groups.
Health needs	A requirement in the population for health intervention or services caused by a health issue.
Health outcomes	The measure of change in the of health status of the population.
Health promotion	Health promotion is the process of enabling people to take control over the negative determinants of health that exert influence on their life and thereby improve their health.
Health status	The level of health of the population or population sub-groups.
Health Survey for England	The Health Survey for England (HSE) comprises a series of annual surveys beginning in 1991. It part of an overall programme of surveys commissioned the Department of Health and designed to provide regular information on various aspects of the nation's health.
Hospital Episode Statistics	Hospital Episode Statistics are data contained in the national statistical data warehouse for England of the care provided by NHS hospitals and for NHS hospital patients treated elsewhere.
Human capital	The stocks of expertise and knowledge of individuals.
Incidence	The incidence of a disease in society corresponds to the number of new cases arising in a given time frame in a specified population group.
Incidence rates	The number of new occurrences of a condition or disease in a population over a period of time.
Index score	Composite score comparing proportion of disease in local population (percentage base) to the proportion in the national population (percentage base).
Internal competition	Internal competition relates to the factors competing for time and attention of an audience, examples include: psychological factors, pleasure, desire, risk taking and addiction.
Modifiable areal unit problem	If data are aggregated into spatial units they are subjected to two effects: Scale effect: different results occur when data are aggregated for different geographical scales Zoning effect: different results occur when data are aggregated for different boundaries.
Morbidity (rates)	The state of living with a disease or condition, therefore it is the number of people living with a condition or disease which decreases quality of life.
Mortality (rates)	The number of deaths occurring for a condition or disease during a specific period of time.
Neighbourhood	Social construct representing an approximate socially homogenous population. In this research the neighbourhood scale is represented by the postcode unit.

Term	Description
Population sub-group	Generic term referring to the identification of different groups of the population where individuals share common social or physical characteristics.
Prevalence	Prevalence rates measures the proportion of a population that are cases at a point in time, it is independent of when the illness began.
Prevalence rates	The measure of a condition or disease in a population at a given point in time.
Risk factors	Characteristics, factors or behaviours for example that increases the likelihood of developing a disease or condition that leads to morbidity.
Service uptake	Service uptake measures the proportion of total invitees who attend a service for example breast screening.
Social capital	Social capital corresponds to the networks, cultural norms , values and supports
Social determinants of health	Some people live longer than others, social determinants of health consider the social influences that result in good or bad health. Such examples include: lifestyle, education, occupation, family structure, access to transport, access to health services, stress, sanitation and exposure to environmental hazards.
Social exchange	Social exchange corresponds to the exchange of social and material resources as a fundamental form of human interaction.
Social facts	Social facts are external to the individual and are general throughout society.
Social interactionism	An approach to technology diffusion that takes into account the users and their environment prior to system development.
Social marketing	Application of conventional marketing approaches together with other techniques that will bring about positive behaviour change to reduce health-harming behaviours.
Spatial autocorrelation	Spatial autocorrelation refers to the extent an occurrence of an event in an areal unit constrains, or makes more probable, the occurrence of an event in a neighbouring areal unit.
Standardised mortality ratios	Standard mortality ratios measure for an area, number of deaths, expressed per 100,000, that would occur in that area if it had the same age structure as the standard population and the local age-specific rates of the area applied.
Weighted Risk Index Scores	Index scores compare the average rate of a variable in a neighbourhood Type compared to the average in the total population. A weighted risk index score is a calculated by weighting the variables fist to ensure that are representative of the distribution of Types in the population. Score of 100 represent the national average.

Table of Contents

Table of Figures

Table of Tables

Chapter 1

Introduction

1 Introduction

1.1 Health inequalities, social marketing, geodemographics and other local neighbourhood indicators

The exploration of health inequalities and their corresponding social determinants has been a topic of research and investigation in the UK for many decades and even centuries. It is a subject area whereby research focuses upon unpacking the causal paths to different diseases and conditions. There are many widely differing determinants on the casual path of a disease or health-harming condition, any of which may influence our predicted health outcomes. These range from individual genes, gender, and age to lifestyles choices, social networks, housing, environment, exposure to immunisation programmes and public policy to name just a few (Marmot et al., 1984, Marmot and Brunner, 2005, Marmot and Wilkinson, 2004, Graham, 2001, Shaw et al., 1999, Bartley, 2004). In the last century the nature of illness has fundamentally changed. Previously, communicable diseases were prominent amongst the primary causes of long-term poor health (morbidity) and death (mortality) (Hicks and Allen, 1999). Since World War 2, in the UK, there has been considerable change. Illness and poor health has become much more closely related to diseases associated with an increasingly affluent society and comfortable living (Choi et al., 2005). Morbidity in the UK today is now closely related to our individual health choices and behaviours such as smoking, obesity, diabetes and alcohol consumption.

It is the associations between illnesses and relative comfort that have led to the rise in recent years of UK government policy to adopt social marketing techniques. Social marketing remains a fairly new subject area for research. It was an idea first conceived in the 1950s and 1960s (Andreasen, 2003; Kotler and Levy, 1969; Weibe, 1951) and is still a reasonably young field with unclear definitions. One widely recognised definition of social marketing is "the systematic application of marketing alongside other concepts and techniques, to achieve specific behavioural goals, for a social good" (National Social Marketing Centre 2007). Social marketing applied to health, is the systematic application of marketing concepts and techniques, to achieve specific behavioural goals, to improve health and reduce inequalities" (French and Blair-Stevens, 2006, page 8). In the traditional sense "to do social marketing" means to apply the principles of marketing theory commonplace in the commercial sector to encourage individuals to consume goods and service, but in this instance it is applied so as to encourage behaviour change related to health harming lifestyles. Principally, social marketing aims to apply conventional marketing theory, with central importance upon product, place, price and promotion, known as the marketing mix (Kotler and Levey, 1969), to social causes to encourage and stimulate behaviour change. With emphasis placed on understanding the product, price, place and promotion, social marketing inevitably lends itself particularly well to geographical segmentation analysis. This in turn leads organically to the requirement to identify differing health needs of differing population sub-groups. It is a process with enormous potential to empower local communities and neighbourhoods to make positive changes to improve their health outcomes and morbidity as identified in the 2004 Public Health White Paper.

To this end, this thesis sets out to develop and explore an alternative framework for measuring and disseminating health outcome indicators associated with health inequality, to be used in social marketing and public health interventions. Geodemographic classifications (small area measures of social-economic and lifestyle conditions) were chosen to facilitate the development of this alternative framework for exploring health outcome indicators. They were chosen because of their successful development and application within the commercial marketing sector as noted by (Batey and Brown, 1995, Birkin and Clarke 1998; Harris et al., 2005; Longley et al., 2001) and it seems natural that they have potential to be extended to the social marketing process for health. They provide enriched detail at the neighbourhood scale (small area equating social similarity) with which differential health outcomes and existing patterns of health outcomes can be explored; they describe distinctive and socially cohesive population sub-groups. This moves the research beyond the traditional 'container' approach of investigation whereby coarse administrative boundaries, known as formal organisational boundaries (Gatrell, 1997) have little relation to the phenomena being measured are used to aggregate health phenomena.

This thesis aims to contribute to the research agenda by developing a geographic and geodemographic framework for exploring inequalities in health outcomes at the functional scale of the local neighbourhood. Fundamental to this approach is the notion that geodemographic classifications reflect homogenous social spaces. New robust indicators of health outcome distributions in population sub-groups will be developed. This is moving beyond the standard aggregations of indicators developed for the formal geographical scales that are created from the nested hierarchical organisation structure that is used to administer budgets and care in the NHS, to consider more efficient and effective discriminators of health outcomes that can be practically implemented in a public health setting and provides a new dimension to this research field.

1.2 Aims and objectives

Key themes considered in the literature focus upon existing research in the fields of health inequalities, geodemographics, social marketing and social capital alongside how they are measured. This leads to the formation of the following overarching aim for this thesis:

> **To develop and explore an alternative framework for measuring and disseminating health outcome indicators of inequality, to be used in social marketing and public health interventions.**

This thesis will present a straightforward yet effective alternative method of exploring the measurement of inequalities in health outcomes, by using geodemographic classifications to develop standardised neighbourhood risk indices (indicators which compare the predicted level of health in a population sub-group compared to the population as a whole). Furthermore, exploration of these inequality gradients through social and geographical space will provide new and innovative insights, alongside greater understanding of local neighbourhoods for health practitioners and lay community members.

The development of an alternative framework for modelling inequalities according to the social similarity of neighbourhoods is justified for the following reasons. Measures based on population similarity move away from the traditional 'container' approach of measurement by zonally coarse administrative unit, as noted by (Harris and Longley (2002) and extends the traditional notion of social measurement to consider social similarity of different populations, attempting to reduce conventional confounding issues that arise from the size of the units. Social measurement analysis is traditionally defined as "social measurements translate observed characteristics of individuals, events, relationships, organizations, societies, etc into symbolic classifications that enable reasoning of a verbal or logical or mathematical nature" (Heise, 2001). Using a social measurement approach to explore how and why population groups behave in certain ways may lead to the identification social health facts, somewhat similar to of the idea of 'social facts' proposed by Durkheim (1895) which leads a discussion on their potential to explore social and geographical space. This is because the way in which a phenomenon is conceived determines the method for which it is analysed (Longley, 2005). Such an approach should provide a useful and systematic approach to the complexity of social classification.

Social marketing and health promoting interventions rely upon a comprehensive understanding of local populations to ensure that their needs and requirements are fully recognised. As identified by Dedman and colleagues (2006, page 2), it is an intelligence led approach, requiring specific and robust data methods for identifying the most relevant population groups. The gathering of social health 'facts' using a geodemographic framework should assist the social marketers in their understanding of:

1. the product: long term good health;
2. the place; which neighbourhoods have the highest risk;
3. the price; reduction in long-term morbidity for changing behaviours patterns and choices;
4. the promotion; which methods should be used to stimulate and encourage behaviour change.

This type of framework, supporting the gathering of detailed local knowledge of the population to ensure services and interventions are specifically tailored, can be considered in the context of the types of support infrastructures for decision-making as proposed by Longley and colleagues in

2005 (Chapter 1). The geodemographic framework presented in this research builds the continuum transforming data facts into information and evidence, which subsequently leads to knowledge and wisdom upon which policies and social marketing strategies can be accepted by decision makers and implemented. Therefore the objectives of the research are three-fold:

1. To develop a framework for extension of geodemographics to health inequalities and outcomes;
2. To assess health variability using this framework in a real world setting and to enable investigation of scalability and validity of the measures developed;
3. To embed the framework in Public Health– and in so doing, transfer research knowledge to professional knowledge.

The geodemographic framework will be evaluated through a series of case studies each reviewing the ability of geodemographics to provide measures that incorporate compositional and contextual influences on health need. This in turn will provide:

a) An enriched picture of local neighbourhood health needs that extends beyond standard deprivation measures;
b) A framework for identifying neighbourhoods at risk;
c) A means of highlighting neighbourhood variation in likely health outcomes and service equality;
d) Proxy indicators of social capital;
e) An intelligence led solution for social marketing campaigns;
f) A means of describing the health trajectories of small localities to explore how lifestyle may be represented using geodemographic indicators;

The case study approach is adopted for a number of reasons. Firstly, using geodemographics the research includes both a social and geographical dimension to the analysis which introduces more complexity. If the framework approach was more theoretically based there is the potential that this might lead to misleading assumptions and analysis. In the case of geodemographics, the case study based approach is the correct one; it is exposing real world problems and challenges that would not be found otherwise. Developing the framework in the context of the different case studies associated with current public policy issues will ensure that knowledge transfer is more successful, because they will present suitable and relevant alternatives to more traditional methods of identifying health variation in the population. In this context the case studies present an inductive method of applying geodemographics for exploring social marketing dilemmas. The case studies present a series of examples which can be used to identify patterns of health variation across different neighbourhoods to identify social facts. Deductive analysis would require a predefined set of patterns which could be explored to determine if the health variations fit, but in this research no prior assumptions are made and as such the variations are unknowns then the deductive approach is not appropriate.

The other reason for adopting the case study approach is that the research was carried out in parallel with a full time Knowledge Transfer Partnership (KTP) with Camden Primary Care Trust (PCT) which was exploring the applications of geodemographics in a public health setting. For this reason the case study approach was deemed a suitable mechanism for transferring knowledge, working with public health staff to observe the practicalities of implementing a geodemographic framework in a PCT.

The resultant framework will be substantiated and the flexibility, scalability and validity of the measures explored. Finally the thesis will explore the challenges and limitations encountered when transferring the acquired research knowledge into professional knowledge through the dissemination of results within a public health department of an Inner London PCT.

1.3 Methodology, analysis and outputs

The measurement of health needs and their associated outcomes forms a fundamental component of evidence-based policy, strategy and delivery of health care services at local scales. These evidence based polices have the potential to reduce overall population gradients of health inequalities which are evident today in UK society (Dorling et al., 2007). By harnessing the power of a spatially enabled geodemographic based framework alongside national health datasets and surveys of population health, the methods developed and explored in this thesis should facilitate the differentiation between local variations in health behaviours across small neighbourhoods. Ultimately this will enable the creation of expected measures/risk indices of health needs for local populations to become commonplace in public health practice.

The research will first set out to define an appropriate local spatial data infrastructure that integrates multiple disparate health datasets into the same functional scale of local geography, the neighbourhood. The framework is then developed by combining operational transactional health data and health survey data to create health risk indices for different neighbourhoods. Two key datasets are used: the annual national Health Survey for England (HSE); and the Hospital Episode Statistics (HES), which is a national database of hospital admissions (a transactional database recording details of all patients who stay at least one night in hospital). These datasets complement each other and these sources facilitate the creation of neighbourhood scale indicators of health behaviours and risk, consequently enabling the prediction of the likelihood of local levels of health need.

Another aim of the thesis is to consider the real world application of these new measures. With the further aim to turn research knowledge into professional knowledge, a final output of the thesis implements the basic principles and methods within the local public health department of one Primary Care Trust (PCT). Camden PCT was chosen because this research was conducted alongside a Knowledge Transfer Partnership between University College London (UCL) and the trust, one of the co-sponsors of this research.

1.4 Thesis structure

This thesis is comprised of 10 chapters, with this the first providing the introduction for the research and the last drawing together the learning outcomes and summarising the main conclusions and areas of further work that would naturally follow. Within these chapters the thesis is comprised of three main sections. Section I explores the research that has gone on before and considers the framework and methodology which will be used in the subsequent analysis. Section II is comprised of the development of the framework through empirical analysis of case studies; the ensuing outcomes are summarised and discussed in the final chapter of this section. It is here that objectives 1 and 2 of the research are appraised. The third and final section (III) is concerned with realising objective three of the research, namely the implementation of the framework in a public health setting. In Section I the most relevant literature is critiqued and the main research objectives identified. The thesis straddles a number of research fields -- health inequalities, social marketing and social capital, geodemographics, and health inequalities measurement – and so the literature review is split between Chapters 2 and 3.

Chapter 2 considers the policy context within which the research objectives and framework reside. It is in this chapter that the meaning and relevance of health inequalities, health outcomes and their consequential social determinants are discussed. The policy framework surrounding the inequalities and the social marketing agenda contextualises the practical applications that this research can contribute to by providing of a robust evidence base for quantifying inequalities and a solid basis for informing social marketing campaigns.

Chapter 3 begins by taking account of the policy situation in the UK before turning to consider existing measurements of health needs, inequalities and expectations. It reviews in detail the various different measures of health outcomes and the traditional deprivation measures that are used a proxy for health inequalities because of the close relationship between wealth and health. It identifies the limitations and research gaps of existing techniques and sets the research agenda for the development of the alternative framework proposed in this thesis. A review of geodemographic classifications is undertaken in this chapter, discussing common methods for their creation, appraising how they have been used in the past, and exploring the ethical considerations that should be evaluated at the outset of the research. In essence this chapter enables the evaluation of the existing measurement techniques against the potential of new alternative techniques, to formulate the potential contributions that an alternative geodemographics framework will add to the health inequalities and social marketing research agenda.

Chapters 4 and 5 consolidates the research literature and presents a detailed synthesis of the research framework and method that binds together this thesis. Chapter 4 describes the three main research objectives that this thesis sets out to explore and evaluate. It documents the evidence from the literature that justifies choice of objectives and suggests how they might provide a new contribution to the research fields. Chapter 5 builds upon the research framework and objectives to outline the research method and processes that will be utilised in realising the research. This chapter also provides the reader with a detailed description of the datasets used as input variables for the framework. Finally an outline of the population characteristics of the main study area is presented in order to introduce the reader to the life laboratory.

The end of Chapter 5 marks the end of the first section of the thesis and signals the beginning of the second. This second section is concerned with developing and implementing a geodemographics framework to explore the first two research objectives proposed using of a series of case studies. Chapter 6 takes the Health Survey for England and creates local neighbourhood indicators of health-harming behaviours linked to smoking, obesity and alcohol. It explores the extent to which the framework can be used to explore differential patterns of these behaviours at the local geographical scale. This chapter focuses upon the exploration of geodemographic and health outcomes to develop measures related to diseases of comfort and health harming behavioural choices. Chapter 7 considers these measures and sets out to substantiate them, and considers the use of operational health datasets to determine how geodemographics can be used for preventative health outcome measurements and to inform existing public health services and social marketing campaigns. In this context this chapter also considers both their validity and scalability of the measures in reference to diabetes risk.

Chapter 8 is the final chapter in this section and acts as a summary and discussion chapter for the review of the analysis developed in the previous chapters of this second section. It has two distinct parts: the first is concerned with the accuracy and validity of such classifications and second discusses issues related to scale. Throughout the case studies presented in this chapter, a number of recurring themes and points of discussion are identified and reiterated. These themes are considered in this chapter, drawing together the analysis into a detailed discussion addressing some of the main points. Discussion surrounds their ability to discriminate those affected by 'diseases of comfort' which is proven throughout Chapter 6 and 7 with patterns of smoking, diabetes, teenage pregnancies to name a few health-harming conditions successfully differentiated, but leads to the need to review how accurate the classifications represent the local neighbourhood populations. This raises the discussion topic of the ability of geodemographics to act as representations of social reality. This chapter also evaluates the effectiveness of geodemographic classifications considering their ability to be effective discriminators. Also up for discussion in this chapter is the notion that geodemographic classifications are a practical and useful alternative to standard deprivation measures. The final discussion point of this chapter contemplates the idea that geodemographic classifications provide a practical mechanism for exploring social space and health status based upon the similarity of neighbourhoods.

Chapter 9 then forms the first chapter of the final section of the thesis. This chapter corresponds to the third and final research objective which is concerned with assessing how research knowledge can be turned into professional knowledge via the dissemination of the framework and research results to public health professionals. The chapter considers existing uptake of spatial knowledge in the NHS and looks at how realistic it is to implement innovative ideas into a hierarchical organisation under constant change and pressure.

Chapter 10 summarises the final concluding remarks that have arisen from conducting this research. Throughout the thesis where areas of further work are identified they are unified in this closing chapter in order to identify and consolidate promising areas of new work that will extend this research into the future. Additionally, the chapter reviews the limitations of the research and where appropriate proposes potentially interesting areas for future development.

Section I

Chapter 2

Literature Review

Part 1: Policy context and theory

2 Literature review

This chapter outlines the results of the literature search which has been used to define the research agenda and objectives of this thesis. The application of geodemographic measures for social marketing is by its very nature multidisciplinary and the research bridges a number of different research fields. The primary purpose of social marketing is to encourage positive behaviour change to reduce the degree of health-harming behaviour in society, the underlying premise of which is to ultimately reduce health inequalities. For these reasons the key research fields explored are: health inequalities, social marketing, geodemographics and the measurement of health needs/outcomes using geographical information systems (GIS). The literature review is divided into two, with the first part reviewing the UK policy context and theoretical framework where this research sits. It explores the methods for how health-harming behaviours, diseases of comfort and their related illnesses have become a number one priority for the NHS, within the health inequalities framework. The chapter reviews the history of health inequalities within the UK and outlines one new UK policy recommendation for reducing health inequalities, defined as social marketing. Within this context the chapter also considers the notion of social capital, identifying its place for aiding behaviour change related to diseases of comfort. The second part of the literature review draws upon the previous research fields and looks at the contribution that geodemographics (small area measure of social, economic and lifestyle conditions) have made to the public sector and the potential they have to assist the measurement of need relating to diseases of comfort and its application to the social marketing domain.

2.1 Literature introduction

In England and Wales over the last century the general patterns in the causes of death have changed significantly. Today the nature of mortality and morbidity in the modern day is very different to that of 100 years ago. Figure 1 provides statistical evidence illustrating these changes. The graph was taken from a 1999 House of Commons statistical report (Hicks and Allen, 1999) and illustrates the most noticeable changes during this time. In 1997 almost two thirds of deaths (69%) were related to neoplasms (all types of cancers) or circulatory diseases (including heart disease). In 1880 as causes of death these diseases were apparently rare. Less than 10% of the population were recorded as dying from these illnesses. In 1880 by far the most common causes of mortality resulted from communicable diseases. One out of three deaths could be linked to infections and parasitic diseases. In 1997 less than 17% of deaths could be attributed to diseases such as tuberculosis, scarlet fever or diphtheria. In society today, diseases related to lifestyle and health choices have become much more prevalent.

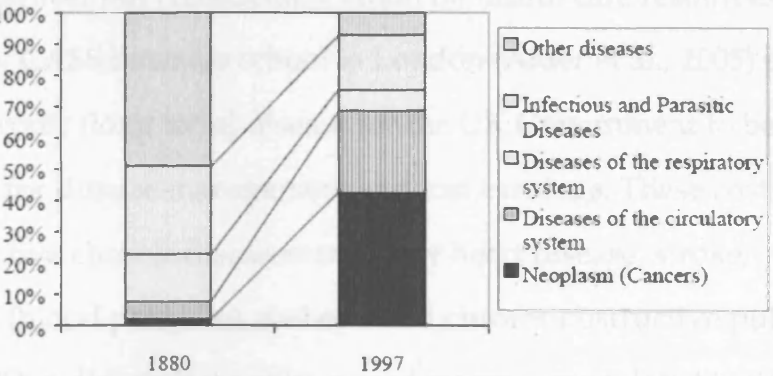


Figure 1: Causes of death in England and Wales: 1880 and 1997

(data source: Hicks and Allen, 1999)

Albeit a simplistic representation and subject to inaccuracies and changes in recording and reporting mechanisms over time, the data in Figure 1 illustrates a key challenge for present day health care policy. Modern day diseases that cause death and long term discomfort are known as chronic diseases, referring to the notion that these diseases persist over extended time periods. Examples of chronic diseases include heart disease, cancer, diabetes and obesity. It is this change in the causes of death (moving from communicable diseases to chronic disease) which has signalled new challenges for public health care practitioners. This is because lifestyle and behaviour are now important due to their ability to increase risk to developing poor health.

For some chronic diseases and health conditions (for example obesity and excessive drinking), their relationship with health damaging lifestyle and behavioural choices has been acknowledged through studies of the social determinants of health. These health damaging behaviours, combined with adverse social and environmental conditions, comprise some of the primary determinants of poor health outcomes in modern England. The considerable influence of health damaging behaviours on chronic disease morbidity and mortality led to coining of the term 'diseases of comfort' by Choi et al. (2005).

Diseases of comfort put considerable strain on health care resources. In 2005 a project run by CASS business school in London (Alder et al., 2005) estimated the cost of chronic (long term) disease for the UK Government to be £12 billion a year for disease management and lost earnings. These costs were estimated for five chronic diseases: coronary heart disease, stroke, hypertension (blood pressure), diabetes and chronic obstructive pulmonary disease (COPD, a disease primarily caused by smoking related behaviours). If preventative health care can lead to a reduction in these types of illnesses there would be considerable cost savings with respect to monetary and morbidity gains.

The rise in the incident rates of these lifestyle diseases in the UK and across the western world cannot be refuted, but they do present many challenges to researchers trying to unpack the complexity of both the population and the determinants of individual health. Genetic inheritances combined with risk exposure to infections and other hazards and the wide variations in the degree of resistance which some people have, together with lifestyle and behavioural choices all influence the individual's health outcomes (Graham, 2001). This complexity has led to a number of models of the social determinants of health to be derived, most famously the one presented by Dahlgren and Whitehead (1991). Two types of models will be discussed in detail in section 2.3.2. Variations in these social determinants of health have resulted in differing health outcomes across the population and their constituent population sub-groups. People living in disadvantaged circumstances are prone to more illness, greater distress, more disability and shorter lives (Benzeval et al., 1995), endemic characteristics of all modern societies both in developing and developed nations. These differences in health outcomes are referred to as health inequalities, which form one of the underlying policy drivers for this research.

Inequalities in health outcomes are a relatively common phenomenon that manifest at distinct spatial and social scales. They occur because of measured observed and predicted variations in health outcomes and status (Dorling, 2007 and Townsend, 1988). In the UK there is a vast array of data and research investigating the link between health inequalities and poverty and/or deprivation. Britain has a long history of health inequalities research, dating back to the 19th century when Victorian philanthropists began to record information relating to deaths alongside the occupation of the deceased. This has led health research, evaluation and analysis, around population attributes related to social class, income, poverty and/or deprivation (Townsend, 1988; Blane et al., 1996; Jarvis and Wardle, 2002; Marmot and Wilkinson, 2004; Rapheal, 2006; Dorling, 2007) Although recently it was noted that health inequalities are not restricted to the poorest or most deprived communities and neighbourhoods, they run right across the social spectrum and different types of people will have varying levels of negative and positive health outcomes (Marmot and Wilkinson, 2004).

Recent policy initiatives driving the health inequalities agenda in the UK and England, place emphasis on the individual taking responsibility for their own health, by making appropriate lifestyle choices and behaviours. At the same time the government has charged the NHS with ensuring its service provisions are efficient and equitable whilst assuring value for money. In an effort to meet policy targets a number of new initiatives have been identified. They aim to target more effectively the populations most at risk of adverse health outcomes associated to different types of lifestyles and behaviour. One such new initiative is linked to social marketing practice.

Social marketing is considered to be an effective mechanism for expediting behaviour change in population groups. Using it as a vehicle for reducing negative health outcomes is a notion growing in popularity. As such it is one application domain for the indicators developed in this thesis. It is believed by policy makers that social marketing techniques, driven by local and national policy, aimed at the appropriate population sub-groups will facilitate positive behaviour change and by corollary reduce health inequalities, improve the efficiency of initiatives and reduce morbidity linked to lifestyles.

In order to develop successful social marketing strategies and improve the efficiency and efficacy of public health initiatives there must be robust and practical methods for understanding populations, for predicting their health outcomes and exploring distributions in inequalities. For this reason the second part of the literature review will conduct a high level survey of existing measurement techniques for predicting health outcomes, status and inequalities and focus in detail on different geodemographic methods and their potential. The learning outcomes gathered from the two parts of the literature review motivate the development of the research objectives and the subsequent chapters and led to the decision to develop an alternative empirical framework with which to explore differing health outcomes to investigating health inequalities for neighbourhoods.

2.2 Policy context

Spurred on by the changes in health outcomes and status linked to the rise in diseases of comfort and the presence of health inequalities across various population sub-groups, the Labour Party commissioned a number of key NHS service reviews led by David Wanless. The first review that was delivered in 2002 was entitled 'Securing our future health: taking a long-term view'. It provided a comprehensive assessment of the resources required to provide high-quality health services in the future. It was based on first catching up, and then keeping up with other developed countries, which had moved ahead of the UK over previous decades before 2000 (Wanless, 2002, Page 9).

This review was the first ever evidence-based assessment of the long-term resource requirements for the NHS. It looked at three different NHS resource cost scenarios (low, medium and high), and identified a number of areas which could be implemented to lower projected overall resource requirements. Ultimately the report highlighted the expected NHS cost differentials that would arise according to its own service productivity and how well the population was engaged in taking responsibility for their personal health outcomes. The report noted that if public health practices were improved there would be, "a substantially larger positive impact on health needs from the focus on health promotion and disease prevention", (Wanless, 2002, page 119).

The least expensive resource cost scenario identified in the review also simultaneously predicted delivery of better health outcomes and was known as the "fully engaged" scenario. It was proposed that this model could facilitate the following: a high level of public engagement, increase in life expectancy beyond current forecasts, improvement in health status, effective and efficient resource allocation and finally a responsive service with high rates of technology uptake. This model was adopted by the government.

Succeeding the NHS resource review was the highly regarded and influential review paper, 'Securing good health for the whole population' (Wanless, 2004). This was a comprehensive independent review again led by Derek Wanless investigating, "prevention and the wider determinants of health in England, and the cost-effectiveness of action that can be taken to improve the health of the whole population and to reduce health inequalities" (Wanless, 2004, page 3). Wanless (2004, page 4) stated that whilst individuals are ultimately responsible for their own health, it needs facilitating by appropriate enabling support structures provided by health professionals and government alike.

Amongst the many review findings was the requirement for the government to develop a more coherent strategy to reduce preventable illness caused by unhealthy behaviour such as smoking and physical inactivity, the diseases of comfort. A central theme of the review noted the role of public health for disease prevention and for encouraging and ensuring the population becomes fully engaged in their health outcomes.

Building on the findings of the initial Wanless report the government published their public health White Paper, 'Choosing health: making health choices easier' (2004). This paper developed the fully-engaged model of health care and set out the principles for supporting and empowering the population to make healthier and informed choices. The White Paper prioritised social marketing as a mechanism for guiding future health promotion efforts directed at achieving behavioural goals (Jesson, 2007). The paradigm shift was announced in Chapter 2 of the White Paper titled "Health in a consumer society", which noted that a wide range of lifestyle choices are marketed to people, although health as a commodity itself has yet to be marketed.

2.3 Health inequalities

It was the significance of existing health inequalities identified by the Wanless report that realigned the priorities of health practitioners and made inequalities centrally important to many government health policies; although as an area of research it is not a new field. Health inequalities are described as the "differences in people's health between geographical areas and between different groups of people" (DoH, 2004, page 37). In England, one of the most infamous health inequalities corresponds to the observable differences in health between population groups living in North England, who have poor health, and the population in South England who are comparatively much healthier at the regional scale of geography (Townsend 1988).

Health inequalities research in the UK is certainly not a new phenomenon; having been on the research and policy agenda for more than a century. In its current guise there are been much corroboration between health inequalities and associated measurements of deprivation or poverty which are used to target health intervention. Evidence for the innumerable health divides is both plentiful and irrefutable, and were identified most notably by the Black Report (Townsend 1988) and reiterated many times since. Within the field of Public Health, the phenomenon of health inequalities has undergone observation for many decades (Marmot et al., 1984, Marmot and Brunner, 2005, Marmot and Wilkinson, 2004, Graham, 2001, Shaw et al., 1999, Bartley, 2004).

Unravelling the complex interactions that contribute to an individual's health outcomes and population inequalities is a challenging task. Individual factors such as gender, age and ethnicity together with lifestyle and behaviour choices and wider issues such as government policy, access to services and the environment all intertwine together to contribute to both population and individual health outcomes, where disparities at the population scale result in health inequalities evident across regions, countries and continents. The term health outcome makes reference to the health status of an individual or population. For example, a person with poor health outcomes has an increased risk of: mortality (death) and morbidity (illness or disability). Health outcomes can be defined more comprehensively as the, "changes in health status (mortality and morbidity) which result from the provision of health (or other) services" (EOHSP, 2007, Online). Variations in health outcomes result in gradients of health inequalities for populations and population sub-groups (Marmot and Brunner, 2005).

Life expectancy is one common measure used to describe diverging health outcomes of different population sub-groups. It will be used here to illustrate the notion of inequalities to the reader. Life expectancy at birth is, "a summary measure of the all cause mortality rates in an area in a given period. It is the average number of years a new-born baby would survive, were he or she to experience the particular area's age-specific mortality rates for that time period throughout his or her life" (Department of Health, 2007).

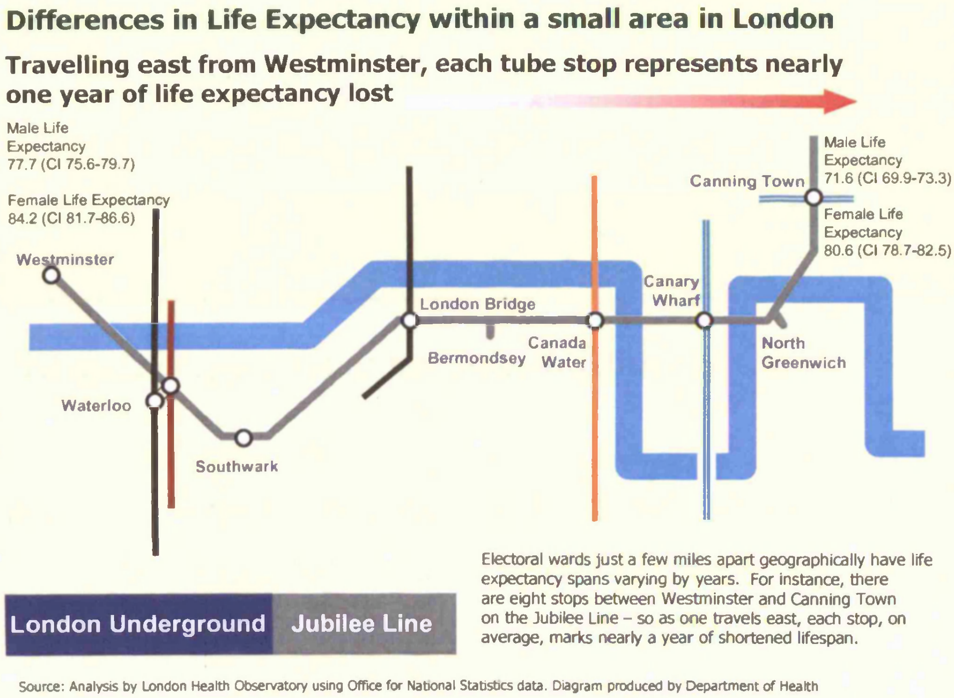


Figure 2: London tube map of life expectancy

(LHO, 2007)

Figure 2 provides one visual example of the geographical disparities in life expectancy across London, UK. The tube map has been used to cartogrametrically illustrate the reduction in life expectancy across a west-east transect of London. If you begin the journey in Westminster (west) female life expectancy is 84.2 years but as you travel east along the Jubilee line to the destination of Canning Town female life expectancy has fallen to 80.6 years. Health inequalities research is interested in understanding why there should be such difference in the population. The case of London has been used for illustrative purposes but diversity in health outcomes are seen across the UK and indeed the world, the figures for life expectancy in the UK are reflected in Table 1 and highlights differences between men and women in Scotland, Wales, England and Northern Ireland.

Life expectancy, 2004-2006

	At birth		At age 65	
	Males	Females	Males	Females
				Years
England	77.2	81.5	17.1	19.9
Wales	76.6	80.9	16.7	19.5
Scotland	74.6	79.6	15.8	18.6
Northern Ireland	76.1	81.0	16.6	19.5
United Kingdom	76.9	81.3	16.9	19.7

Table 1: Life expectancy in the UK, 2004 to 2006

(source: ONS, 2008)

2.3.1 Historical appraisal of health inequalities in UK

The historical evidence for the presence of health inequalities is well documented. In the UK, William Farr, a nineteenth century epidemiologist, established a system for recording all deaths in England and Wales. The system recorded the occupation for each dead person and their corresponding cause of death, enabling comparison of mortality rates by occupation. Farr's work used the collected statistical data to test social hypotheses.

During the time Farr was working and developing the mortality records, cholera was prevalent across London. Farr's meticulous records of mortality were pivotal to the first application of spatial analysis and health. John Snow, a London physician, suggested that cholera was an air-borne disease which he presented in 1855 in a paper titled, 'On the Mode of Communication of Cholera'. During an ensuing cholera outbreak Farr was able to provide John Snow with epidemiological data from his mortality recording system. Snow painstakingly mapped the locations of people where deaths that could be attributed to cholera together with the location of water pumps in the neighbourhood (Lang, 2000). He was able to associate the source of the outbreak to a water pump in the centre of a cluster of deaths.

At this time in England demand for state intervention in public health began. The philosopher Jeremy Bentham promoted the, 'sanitary reform movement' that began in the 1820s and 1830s. He placed emphasis on the notion that the state should bear some responsibility for the health of its people (Farmer and Lawrenson, 2004). Edwin Chadwick carried on these ideas promoted by Bentham and Farr by studying the sanitary conditions of the labouring population of Great Britain in 1842. He highlighted the economic cost of an unhealthy workforce. Following his report came the first Public Health Act in 1848 and subsequent acts in 1875 (Disraeli's Public Health Act), 1936 and 1961.

More recently a monumental report was published in 1980 following the appointment of Sir Douglas Black, as chair, who reviewed the evidence of inequalities in health across modern England. The Black report highlighted health inequalities in different populations had been growing since the 1950s (Townsend et al., 1988). The authors illustrated the existence of large differentials in mortality (death) and morbidity (sickness/illness/disease discomfort) that favoured those people associated to higher social classes (Shaw et al., 1999).

The authors of the Black report note four theoretical explanations for the link between health and inequality. Firstly explanations linked to artefact were suggested because health outcomes were artificially derived variables that attempt to measure social phenomena, and therefore the results may be unrepresentative. Secondly, theories of natural /social selection were proposed, whereby the occupational class structure was seen as a filter of human beings. Thirdly, materialistic or structuralist explanations were used to emphasise the role of economic and associated socio-structural factors in the distribution of health. The final explanation turned to the difference in cultural or behavioural patterns, an explanation for inequality of considerable importance in today's policy climate.

The recommendations in the Black report were rejected by the government of the day. They were deemed too costly and only 260 copies of the original report were printed. As a result health inequalities continued to grow in the UK through to the end of the last millennium and into the new one. In unison with the increase in health inequalities there was an exponential growth in the research on health inequalities and deprivation; a simple search on PubMed for the term "UK health inequalities" between 1997 to 2005 returned 2955 articles, whereas between 1979 and 1996 only 514 articles were listed, though of course more online articles are available for more recent years. Alongside the publication of the Black report two significant research reports highlighted the status of health inequalities in the UK, both of which pre-empted and were precursors to the 2002 Wanless review. The Whitehall longitudinal study and the Acheson report were pivotal pieces of work that outlined the disparities of health across different social groups in the UK.

The Whitehall study was a longitudinal study of 18,000 men working in the UK Civil Service during the 1970's, with the published in the early 1980's. The study's original purpose was not to explore health inequalities but rather to investigate cardio-respiratory and diabetes related diseases and their associated precursors. This is because the exploration of health inequalities was not an initial objective of the study. At the beginning of the 1970s when the research was proposed and first instigated, health inequalities between different social classes were not high on the research or political agendas.

At that time, researchers believed key differences in health highlighted more affluent people suffered from heart disease than poor people who were subjected to diseases related to material deprivation (Marmot and Brunner, 2005). The Whitehall study inadvertently disproved this theory. The original study discovered that men in the lowest employment grades were more likely to die prematurely than men in the highest grades. Writing in 1984, Marmot and his colleagues discovered the inverse association between grade (level) of employment and mortality from Coronary Heart Disease (CHD). Men in the lowest grade had a mortality rate three times higher than men in the highest employment grade.

In 1997 Sir Donald Acheson was asked by the government to lead an independent inquiry into health inequalities. The inquiry reviewed and summarised the status and evidence of health inequalities in England and identified priority areas for the development of policies to reduce them. The inquiry found that, "Inequalities by socioeconomic group, ethnic group and gender can be demonstrated across a wide range of measures of health and the determinants of health" (Department of Health, 1998), page 17). Reducing the widening gap in health outcomes experienced by the UK population has become a primary objective of the present government. To understand how health inequalities can be reduced it is necessary to explore the methods for their manifestation in society and then consider how social marketing can act as a suitable vehicle to incite behaviour change. For this reason the next section reviews the different models of health inequalities presented in UK research. This will provide an understanding of the complex challenges social marketers face when needing to consider behaviour change.

2.3.2 Models of health inequalities

The three reports described in the previous section; the Black Report, The Whitehall Study and the Acheson Review, laid the foundations for the development of four accepted models of health outcomes and their related determinants (Bartley, 2004).

The first model explores the influence of the presence or absence of material wealth and its associated level of deprivation. Material factors such as poor housing, occupational hazards, unemployment or environmental factors such as exposure to air pollution for example will impact negatively on health outcomes, although often it is cited that people have no control over these variables and cannot help but be exposed to them (Shaw et al., 1999). A number of studies in the 1980s and 1990s investigated the use of different census variables acting as proxy measures of material deprivation: Townsend Index of Deprivation (Townsend et al., 1988); Jarman Index of underprivileged areas (Jarman, 1983, 1984); Carstairs' Index (Carstairs and Morris, 1989, 1991); Breadline Britain (Gordon and Pantazis, 1997), and more recently the Index of Multiple Deprivation (IMD) (DCLG, 2004). These measures have been criticised in the literature because they are constrained by aggregating data to administrative units, creating containers of deprivation (Harris and Longley, 2002) and they are discussed in detail in the following literature chapter (Chapter 3).

The second model emphasises the importance of cultural behaviour on health outcomes, and was one of the central theses of the 2004 Wanless report. It considers how social norms and our social interactions and behaviours can shape and influence health outcomes. This model of inequalities has become incorporated into key policy initiatives. The Public Health White paper, "Choosing health: making healthy choices easier", recognises the need to strengthen and empower communities enabling them to make healthy choices (Department of Health, 2004). Interest surrounds the link between certain behavioural patterns and the underlying health inequalities that ensue. Twigg and colleagues (2000) noted the central importance of health-related behaviours to health promotion and the promotion of enhanced population health, but despite this, debate still surrounds whether health-related behaviours should be seen as a cause of health inequalities or the outcome of differences in the material circumstances between socio-economic groups (Shaw et al., 1999).

The third influential category of health inequalities is more contemporary, and has grown since the 1980s (Marmot et al., 1984, Berkman and Kawahci, 2000). It is commonly known as the 'psycho-social' model of health inequalities and refers to the psychological effects of the experience either of the work, home or low social status. In the United Kingdom, Professor Michael Marmot has led the research within the field of psychosocial health inequalities, with the renowned Whitehall Study I and more recently Whitehall Study II (Marmot and Brunner, 2005).

The final model describes the influence of the life-course on health inequalities. The rationale behind the life-course explanations of health inequalities incorporates the concepts of timelines. Incorporating the path from childhood, to adulthood and finally to old age. The three timeframes assist in the understanding of the processes that contribute to health inequalities through socio-economic disadvantage over an individual's lifetime (Graham, 2001).

Whilst each of these models or explanations of health inequalities are formidable in their own right, they do not provide sufficient explanation of inequalities in isolation. For this reason attempts have been made to produce an overall model of the factors influencing health, generally referred to as social determinants of health. The most widely referenced model of health is that produced by Dahlgren and Whitehead(1991).

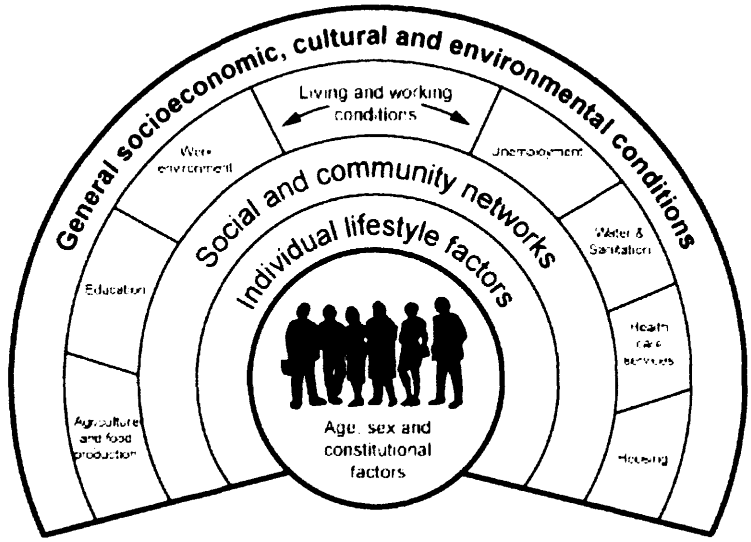


Figure 3: Model of health determinants

(Dahlgren and Whitehead, 1991)

This overarching model of health inequalities incorporates the four key explanations of health inequalities. It clearly illustrates the vast number of variables ranging from the individual to the group level that all contribute to the health of individuals. Although this model is widely accepted by health professionals, there are some limitations to its application. The model fails to highlight the intricate network and spheres of influence that each of the factors wields upon health outcomes and it presents an individual central approach to exploring health inequalities.

Most health inequalities research and many health improvement strategies focus on two of the circles illustrated above: the inner circle investigates biological factors associated to the health of the individual and is encompassed by individual lifestyle variables such as diet, exercise and smoking. Emphasis in recent research and government campaigns have focused on targeting individuals, but the efficacy of this approach is disputed. It is more effective to focus on social systems not just individuals (Blane et al., 1996).

The main premise of public health is to improve the health of populations, thus perhaps it is more appropriate to consider a model of health determinants that is population centric. Figure 4 illustrates the array of variables that may be considered when modelling the health of populations and individuals. The model presents a population centred view of health inequalities. It was developed by Friedman et al. (2002) and is in use by the Centre for Disease Control in the United States.

Figure 4: Model of variables influencing population health

(Friedman et al., 2002)

It is essential to have a comprehensive and coherent grasp of the different variables that interact to influence the health of a population. The focal point of the model is the overall health of the population of interest. The population's health in this model is represented by a number of significant variables or measures which include; the measure of disease, well-being and health status each of which are influenced by the surrounding variables. Aggregate measures within the model are attributes of the community, derived from individuals whilst ecological measures are not derived from individual data and represent contextual data.

One key domain in this population centric model is the influence of the social attributes of the community. Social attributes include understanding networks, cohesion, influence, support and social change. Evidence in the literature links informal social networks, social activities and participation in organisations with better health chances (Cattell, 2001) through social capital which is discussed at the end of the next section.

These social determinants models of health inequalities are useful to have in mind when reading about social marketing because they will influence considerably both the development and overall success of a campaign and how it is implemented. They are also important when considering the development of health outcome measures because the model used will inform the type of metrics developed.

2.4 Social marketing

Health inequalities research is concerned with exploring and explaining population differences in health status and outcomes, which as the social determinant models often showed arise from health-harming lifestyle choices. A recent emergence in public policy is the increasingly important role that social marketing will have of stimulating behaviour change. A natural consequence of this will be the reduction of levels of morbidity in society and rising health status across all population groups, reducing gradients and absolute differences in health inequalities.

The situational context of current policy with its emphasis upon empowering the population to take responsibility for their own health indicates a paradigm shift in public health activities to encourage positive lifestyle and behaviour changes. The consequence of reducing negative lifestyle choices should ultimately achieve the policy goals of reducing health inequalities amongst the population.

Social Marketing is one mechanism for expediting behaviour change in the population which is growing in popularity. One widely recognised definition of social marketing is "the systematic application of marketing alongside other concepts and techniques, to achieve specific behavioural goals, for a social good" (National Social Marketing Centre, 2007). In the traditional sense "to do social marketing" means to apply the principles of marketing theory commonplace in the commercial sector to encourage individuals to consume goods and services, but in this instance it is applied so as to encourage behaviour change related to health-harming lifestyles. The indicators developed in this thesis set out to inform this process, by identifying relevant sections of the population that would benefit from such initiatives.

It is believed by policy makers that through the application of appropriate social marketing techniques, driven by local and national policy, aimed at the appropriate population sub-groups, positive behaviour change may be realised and by corollary a reduction of health inequalities achieved. Due to the importance of social marketing as an application area for the indicators developed in this thesis this next section briefly discusses what it means be a social marketer and describes the existing framework for conducting such activities.

2.4.1 Defining social marketing

There has been and still remains some debate surrounding the precise meaning of the term social marketing, which is briefly discussed in the next section, but in essence it refers to the application of commercial marketing techniques for public good. In this framework public good is an ethical notion applied to policy and decision making, where it refers to a mutually beneficial outcome for society and its population.

Andreasen writing in 2003 traces routes of social marketing to the American sociologist Weibe. In 1951, Wiebes' well cited paper remarked upon the effectiveness of radio and TV advertising being so great that it should also be used to shape behaviour and habit patterns in areas of social responsibility and participation (Weibe, 1951, page 679). In essence Weibe raised the question, "Why can't you sell brotherhood like you can sell soap?" using brotherhood as a reference for social causes. He observed the root of advertising as moving people into interaction with social mechanisms with respect to five motivating factors: force, direction, adequacy and compatibility, mechanism, and distance.

The force refers to the predisposition of the audience and the motivation provided by the communication medium. The direction tells the audience where it can consummate its motivation, whilst the mechanism directly refers to implementation. If the advertising is adequate and compatible with respect to the audience then it will successfully enable the goal behaviour to be achieved. The fifth and final factor corresponds to the distance the audience member believes must be travelled to reach the goal behaviour. This does not necessarily mean spatial distance, but is most often related to energy, effort and motivation. With all of these factors considered Weibe concludes that the seller of a social objective would be in a position, "comparable to that of a commercial sponsor" (Weibe, 1951, page 691).

This paper had two main outcomes. Firstly it led to the thinking that the harnessing of marketing principles and techniques for social objectives could prove beneficial for the public. Secondly the five factors advertising social objections provided the foundations for the conceptual social marketing framework.

In 1969 Kotler and Levy released their seminal paper entitled, "Broadening the concept of marketing". In this paper the authors' contention was that, "marketing is a pervasive societal activity that goes considerably beyond the selling of toothpaste, soap, and steel. ...Political contests remind us that candidates are marketed as well as soap..... No attempt is made to examine whether the principles of "good" marketing in traditional product areas are transferable to the marketing of services, persons, and ideas" (Kotler and Levy, 1969, page 10). The conclusion of the paper is that all organisations: public, private and non-profit have the potential to benefit from traditional business marketing tools.

Two years later in 1971, Kotler and his colleague published another article, exploring the application of social marketing as an approach to planned social change. This was based on their belief that social marketing provided a promising framework for planning and implementing social change (page 3). Building upon previous work and by using the five factors of advertising presented by Weibe, they presented the first formal definition of social marketing. Defining it as, "the design, implementation, and control of programs calculated to influence the acceptability of social ideas and involving considerations of product planning, pricing, communication, distribution and marketing research" (Kotler and Zaltman, 1971, page 4).

A number of authors (Hastings and Haywood, 1991, 1994, reiterated by Andreasen, 2003) highlighted the difficulties with this and subsequent versions of this definition. Andreasen (2003) noted that it proved troublesome for health professionals to understand because many health practitioners saw little difference between this and existing work they carried out.

Thus in 1994 (and 2003) Andreasen proposed a new definition of social marketing. "Social marketing is the application of commercial marketing technologies to the analysis, planning, execution, and evaluation of programs designed to influence the voluntary behaviour of target audiences in order to improve their personal welfare and that of the society of which they are a part" (2003, page 296). This definition ensures that there is no confusion with regards to education and attitudes but states specifically that its main goal is to influence behaviour change.

Hastings and Saren (2003), UK academics conducting research within field of social marketing, largely agreed with Andreasen's redefinition of the term social marketing. But to further the field they proposed that future application of social marketing should incorporate critical marketing and social exchange theory within the agenda to further its efficiency and application. They observed (Hastings and Saren 2003, page 306) that in the UK today, social marketing theory and practice is developing towards more complex and ambitious modes of analysis and understanding. The social marketing definition most widely adopted in the UK, and used in practice, is the more recent one provided by Andreasen (1994, 2003). The importance of social marketing re-emerged in late 1990s as the result of its incorporation into Government Strategy documents.

As identified in section 2.2 the current public health agenda in the UK was laid out in the 2004 White Paper: Choosing health; making healthy choices easier. The driving force behind this strategy was a cross-government drive to reduce poor health outcomes and inequalities, and by corollary improve health and prevent disease. One of the main objectives of the white paper was to enable people to take responsibility for their own health and through provision of support for these choices health outcomes should have more favourable conclusions.

The White Paper discussed the healthy choices and good health in the sense of positive commodities. Paragraph 8, page 20 (DOH, 2004) discusses the requirement to create a demand for healthy choices by noting that,

> *"...A wide range of lifestyle choices are marketed to people, but health itself has not been marketed. Promoting health on the principles that commercial markets use – making it something people aspire to and making healthy choices enjoyable and convenient – will create a stronger demand for health and in turn influence industry to take more account of broader health issues in what they produce".*

This clearly aligns social marketing as a priority in modern health care. As observed by Hastings and McDermott (2006) the new public health agenda primarily driven by government policy has led to the formation of the National Social Marketing Centre of Excellence (NSMC) and the commissioning of the first major review of social marketing to be led by the National Consumer Council (NCC). This new organisation has adapted the definition presented by Andreasen (1994, 2003) and specifically quotes the definition of social marketing as defined by French and Blair-Stevens (2006, page 8). It is the:

> *"the systematic application of marketing, alongside other concepts and techniques, to achieve specific behavioural goals for a social good". The centre also provides a specific definition of social marketing with respect to health,*

> *"the systematic application of marketing concepts and techniques, to achieve specific behavioural goals, to improve health and reduce inequalities"* (French and Blair-Stevens, 2006, page 8).

The health specific definition of social marketing highlights some of the key objectives of the public health white paper (DoH, 2003) and arose from the strategic review of social marketing for health promotion in England, conducted by National Consumer Council (NCC) in 2005. This review acknowledged the potential social marketing techniques had to inform the development and implementation of future health promotion activities. The review identified five key findings:

1. Social marketing can significantly improve impact and effectiveness when applied systematically;
2. There is potential to use available resources and mobilise assets more effectively;
3. Current approaches are unlikely to deliver the required policy goals. Leadership and effective co-ordination are keys to success;
4. Social marketing capacity and capability across the wider public health system is currently underdeveloped.
5. The importance of integrating effective research and evaluation into the development of programmes and campaigns to maximise its value.

As documented by French and Blair-Steven (2006, page 29),"In June 2006 the government accepted the recommendation of the review that social marketing should be used to guide all future health promotion efforts directed at achieving behavioural goals, and it has begun to adopt a national approach to systematically applying social marketing principles to guide its efforts".

2.4.2 Fundamental principles of social marketing

Social Marketing in its current guise has been growing in the UK since the late 1990s. The strategic review of social marketing in 2006 in the UK health sector assisted in highlighting its untapped potential with application to changing health behaviours. The review ultimately led social marketing to be placed firmly on the policy agenda, because of its potential to assist in reducing health inequalities. So whilst we previously discussed the importance of social marketing and its historical routes, this next section explores what it means to actually do social marketing.

In the traditional commercial marketing framework, marketers are primarily interested in four variables: product, price, place and promotion. Indeed, these variables are widely recited in both marketing and social marketing literature and are known colloquially as the four ps of marketing, an expression attributed to McCarthy (1968) by Kotler and Zaltman (1971). In future sections of this thesis the four ps are considered within the context of a geodemographic framework for social marketing.

Kotler and Levey (1969), whilst exploring the potential application of marketing for public good (to influence behaviour such that positive outcomes are beneficial for society), identified a number of different types of products, shown in Table 2.

Table 2: The 5 types of products in traditional marketing

(after Kotler and Levey, 1969)

The table of products Table 2 whilst recognising a number of different 'products' fails to acknowledge the key purpose of social marketing according to its most current definition. In today's world of social marketing there is no product in the sense of a tangible item, the main 'product' for sale today is behaviour change. With regard to health this means transforming health-harming behaviours of population sub-groups into health-improving behaviours.

Kotler and colleagues (Kotler et al., 2002, page 5) identified four types of objectives social marketers wish their audience to achieve: accept a new behaviour, reject a potential behaviour, modify a current behaviour and abandon an old behaviour. With this in mind it is worth exploring the model of social marketing adopted in the UK. The four ps of marketing provide a good starting point, but in reality they are a little restrictive and it is pertinent to adopt a broader approach for defining a social marketing model for health.

The social marketing model proposed by French and Blair-Stevens (2006) is known as the customer triangle model of social marketing (see Figure 5). It encompasses 7 key principles and concepts that are at the heart of all social marketing campaigns and activities.

Figure 5: The customer triangle model of social marketing

(source: French and Blair-Stevens 2006)

Within this consumer triangle the consumers (patients or clients) are placed in the centre. This is because they represent the core motivation for implementing social marketing. The principles suggest that in order to successfully reach the core consumer the surrounding 7 principles (Figure 5) should first be identified and then implemented. The principles are listed as follows: the consumer, insight, exchange, competition, behaviour and behavioural goals, intervention and marketing mix and finally audience segmentation.

Working from the centre of the triangle outwards the 3 principles to be followed are: insight, exchange and competition. The NSMC define insight as the,

> *"actionable insights that are able to provide a practical steer for the selection and development of interventions. To develop such insight means moving beyond traditional information and intelligence (e.g. demographic or epidemiological data) to looking much more closely at why people behave in the ways that they do, to exploring their motivations"* (NSMC, 2006, page 12).

It seems appropriate that geodemographics might have the potential to direct understanding of local neighbourhood insight.

The principal of exchange is crucial to successful social marketing (and commercial marketing). Often the consumer has to invest time, effort or money, so it is essential that perceived benefits equal or exceed the perceived costs (Bagozzi, 1975). It is these costs which equate to the price component of the 4 ps of marketing. "The price of a social marketing product is the cost that the target market associates with adopting the new behaviour" (Kotler et al., 2003, page 217).

As well as being important to social marketing, exchange theory is at the heart of commercial marketing (Rothschild, 1999). "Exchange theory assumes we are need-directed beings with a natural inclination to try and improve our lot" (Hastings and Saren, 2003, page 6). For exchange theory to work social marketers must ensure the "exchange involves the transfer of tangible or intangible items between two or more social actors" (Hastings and Saren, 2003, page 6). The social actors in this instance are the consumer and the social marketer.

Whilst the NSMC identify social exchange as a key requirement in social marketing, its importance has not always been recognised or properly validated for 3 reasons (Hastings and Saren, 2003). Firstly, some social marketers dispute the need for two-way exchange and believe mutual exchange contradicts the altruistic driver behind social marketing. Secondly for successful exchange all parties involved should be capable of communication and delivery, which may not always be the case in social marketing exchanges. Finally successful exchange must deliver clear benefits. But these customer benefits in social marketing are often perceived as ambiguous. They request the customer to change some immediate pleasurable health adverse behaviour to a new behaviour which has some vague probabilistic positive health outcome, that may or may not be realised some time in the future, but certainly not in the immediate future (Rothschild 1999). Social marketing is focused on long-term goals as opposed to the short-term instantly gratifying rewards of commercial marketing – the hare and the tortoise.

Alongside insight and exchange, the customer triangle for social marketing recognises the importance of competition. The concept of competition is, "to examine all the factors that compete for people's attention and willingness or ability to adopt a desired behaviour" (NSMC, 2006, page 13). Both in social and commercial marketing identifying the competition is crucial, but its nature and composition are dissimilar.

Commercial marketers are interested in achieving sustainable competitive advantage and use a number of mechanisms in order to achieve this. For example they often define the competition and analyse it using the views of the consumer (Hastings, 2003). For a commercial organisation competition is considered to be, "Other organisations offering similar goods and services or services that satisfy similar consumer needs" (Kotler et al., 2002, page 10).

In contrast social marketers are in the business of selling a behaviour change so competition is characterised differently. In social marketing competition is most often the, "current or preferred behaviour of the target market" (Kotler et al., 2002, page 10). What is meant by target market will be discussed in more detail later in this section, but refers to the section of the population trying to be reached. Recognising the preferred behaviour as competition is crucial.

Rothschild (1999) stresses the importance of recognising competition within social marketing objectives. "Free choice, apathy, and inertia are powerful competitive forces that often are ignored… there is always competition. For every choice there is an alternative: to be or not to be, to binge drink or drink in moderation"(Rothschild, 1999, page 28). The nature and importance of competition in social marketing is well documented in the literature (Rothschild, 1999, Hastings, 2003 and Clay Wayman et al., 2007).

The NSMC identify two types of competition that impact negatively on social marketing strategies: external competition and internal competition. External competition corresponds to those organisations that deliberately promote negative health impacting behaviour. An example of an external competitor to an existing social marketing campaign would be British American Tobacco targeting smokers who are trying to quit. External competition is not directly associated to the customer whereas internal competition is associated directly to the customer. It relates to the, "power of pleasure, enjoyment, risk taking, habit and addiction that can directly affect a person's behaviour" (NSMC, 2006, page 13).

Figure 6: The customer triangle model of social marketing

(French and Blair-Stevens 2006)

Returning to the customer triangle we now move onto the principles documented around the perimeter of the triangle (highlighted in red in Figure 6): clear behaviour goals, understanding the customer through consumer research and being theory-based and informed. Achieving behaviour change in population groups is a challenging, complex and comprehensive process. Social marketing campaigns must be designed and planned with a specific behaviour objective(s) in mind (Kotler et al, 2002, page 143). Furthermore in the current policy climate, behaviour changes must be measurable and for this reason behaviour objectives are categorised into two types; knowledge objectives or belief objectives. Often consumers need to have both knowledge (information and facts and/or a belief (values, opinions or attitudes) to be empowered to change behaviour.

Knowledge objectives such as statistics and facts relating to the behaviour change will engage customers by motivating them, and belief objectives ensure the social marketer identifies the positive and the problematic behaviours to understand the relationship between them and identifies patterns and trends over time and the influences of them (French and Blair-Stevens 2006). Understanding the knowledge and belief objectives for a behavioural change campaign will provide direction for developing appropriate strategies. French and Blair-Stevens (2006) also note that a successful campaign draws upon theory from a diverse set of professional disciplines. Theoretical underpinnings are essential but should not be applied like a whitewash to all campaigns but must be appropriately selected.

Another guiding principle of social marketing is known as marketing mix. In commercial marketing this is a concept we visited previously as the four p's of marketing: product, price, place and promotion. "The product should be positioned to appeal to the desires of the target market to improve their health or prevent injuries more effectively than the competing behaviour the target market is currently practicing" (Kotler et al., 2002, page 7). To achieve the identified social marketing goal a variety of interventions can be strategically adopted.

Social marketing begins and ends with a focus on the individual within their social context (French and Blair-Stevens 2006). Understanding the consumer is critical to success. Like the commercial branch of marketing local customer knowledge is gathered to gain insight. This ensures a bottom-up approach driven by the health needs of the community and avoids top-down strategically driven campaigns that are ignorant of the customer. Before customer research can be conducted it is first necessary to identify the target audience of the campaign, a critical step for successful campaigns. The theme of this is central to the thesis which leads us to the seventh principal of the social marketing triangle not yet discussed is the necessity to develop a segment approach.

Because social marketing represents a paradigm shift in policy, it is unusual for health practioners to actively conduct targeted campaigns and the use of segmentation methods are rare, with the exception of identifying ethnic groups using census data. As described by Kotler and colleagues selecting the appropriate target audience requires the identification of "a set of buyers sharing common needs or characteristics that the company decides to serve" (Kotler, 2002, page 116). They go on to suggest a 3 step process for selecting target markets: Segment the market, evaluate the segments and choose one or more for targeting. There are a number of benefits that market segmentation can add to the social marketing process. Firstly it provides a mechanism for tailoring interventions by need. Secondly the effectiveness of an intervention is increased. Thirdly resource allocation is directed which increases the likelihood of more successful results. Fourthly, through understanding the target market, social marketers are better able to develop evidence based strategies.

Of the social marketing principles, the step of selecting the target audience is crucial; who are the people with the greatest need? How do you understand their intrinsic motivations? Identification of the most relevant target audiences and ensuring interventions are suitable for their health needs will enable inroads to be made on tackling existing health inequalities in the UK. This review has highlighted the need to implement policy suitable indicators incorporating geographical and population differences in potential health needs and inequalities.

2.4.3 Social capital

The previous section highlighted that since the 1960s there has been a paradigm shift in the conceptual boundaries of marketing (Glenane-Antonaidis et al., 2003). Researchers began to consider the application of marketing principles within a social setting, ameliorating negative behaviour change to benefit the public and improve health outcomes. But the marketing of behaviour changes does present difficult challenges to health professionals: as noted previously the delayed arrival of the rewards, the amount and type of competition, the form of social exchange required and cultural and social norms of society all contribute to reinforcing established behaviour patterns.

Some of these inhibitors to behaviour change are brought together by the notion of social capital. Here social exchange, cohesion, networks and relationships are considered in the context of a type of capital, a set of resources relating to the society. The last 25 years saw an explosion in the growth of research surrounding social capital, its theoretical framework and its impact on society, communities and individuals. The recent growth is the reason the social capital framework is considered to be the most popular export of sociological theory in recent time (Portes, 1998). It encapsulates the belief that involvement and participation will lead to positive consequences both for the individual and the community.

Most widely known is the work of Putman who explored America's declining social capital by investigating the disengagement of Americans from traditional organised civic associations. Writing in the *Journal of Democracy* Putman (1996) explored America's declining public participation with relation to civic actions such as voter turnout, active parent-teacher associations and the decline in organised bowling leagues despite the fact that more Americans bowl now than ever before. During the two decades between 1970 and 1990 voter turnout in the American political elections declined by almost a quarter. He said this represented a disengagement from community affairs by Americans. Putman's perspective on social capital concentrates on the social norms and values of communities and is mostly applied at that level of abstraction. Whilst this is one of the most frequently cited pieces of literature, it is not the only perspective. Alternative views on social capital will be explored in the following sections.

2.4.4 Theoretical roots of social capital

Putman (1996) attributes the modern social capital framework to James S Coleman (within the American group of sociologists), but the roots of contemporary social capital theory can be traced to classical sociologists such as Durkheim, Marx, Weber and Simmel (from Baum and Ziersch, 2003; Woolcock, 1998). Whilst it is beyond the scope of this literature review to appraise in detail classical sociology, it is relevant to focus on some important and relevant aspects; positive solidarity, social exchange theory and social consciousness.

A first key figure is Durkheim, a French sociologist (1858 – 1917) whose work on the linkage between society and health made a considerable contribution to this research field. He published four key works in his lifetime; "The division of labour in society", "The rules of sociological method", "Suicide" and "The elementary forms of the religious life". Durkheim's work, "The Division of Labour" proposed that society is held together by the division of labour. Individuals are dependent upon each other because they each specialise in different types of work. This leads to the notion of positive solidarity that simultaneously binds individuals to society and creates a dependency of the individual upon society (Durkheim, 1964). This implies a positive collectiveness of individuals in society which influences the way in which people interact with each other.

In another of Durkheim's works he investigated the social causes that influenced the suicide rate (proportion of suicides per total population). He theorised that suicide rates were essentially the result of the level of social integration of different groups and their participation in organised religion (Jones, 1986). A key objective of Durkheim's research into suicide was the explanation of individual mortality as a function of social dynamics, although the research was subject to the ecological fallacy (see section 3.5.4). As Berkman and Glass (2000) identify, this publication essentially provided the building blocks for understanding the role of social integration and health outcomes. Portes (1998) identified that Durkheim's emphasis on group life was an antidote to anomie and self destruction and represents the notion of group participation.

Ritxer writing in 2000 suggested the essence of Weber's work, "Economy and Society" explored the theory of social actions and how individuals behave when they encounter each other. It introduced the three components of social action: subjective meaning, social relationships and stable content. His work suggested that when individuals respond to one another the responses have expected and usual patterns which enable the development of social relationships

Each of the classical sociologists highlighted the need to investigate the relations between society and individuals and their associated activities to understand the related impact and behaviour. The theories of social consciousness, social exchange and solidarity are linked to the classical sociologists of the eighteenth and nineteenth centuries and are rooted in economics, sociology and politics. They form the building blocks for modern social capital theory.

There are two main theorists who are responsible for the modern social capital framework; Bourdieu and Coleman. Bourdieu focused on the benefits that individuals acquire through participation in civic organisations proposing that individuals profit from group membership (Bourdieu, 1984). His work identified two elements of social capital theory. The first element suggested that social relationships and networks enable individuals to claim access to resources possessed by their associates and the second element emphasised the quality and amount of resources available (Portes 1998). His work approaches social capital from a network viewpoint and places importance on the individual within the network, exploring the connections between people.

In contrast Coleman (1988, 1990) theorises social capital from the viewpoint of a group or community and places less stress on the social network per se. Coleman explores the role of social capital to create human capital building upon the concepts of social action and rational action developed by Weber (Coleman, 1998). He notes that social capital is derived from the changes in relations among persons that facilitate action and places emphasis on group outcomes.

2.4.5 Definitions of social capital

In the work of Bourdieu the term social capital is linked to the social network paradigm with one central difference. While social capital corresponds to the actual features of social structures that enable the actions and behaviours of individuals (actors), by comparison social network analysis is concerned with the linkages both of and between these individuals.

For Bourdieu (1968, page 248) "social capital is the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance and recognition". In contrast Coleman followed later by Putman considered social capital to be the combined value of relations of individuals, communities or societies. It is described by Putman (1996, page 66) as the features of "social organization such as networks, norms and social trust that facilitate coordination and cooperation for mutual benefit. From this viewpoint social capital enhances the benefits of investment in physical and human capital". Social capital is "a variety of entities with two elements in common: they all consist of some aspect of social structures, and they facilitate certain action of actors – whether persons or corporate actors – within the structure" (Coleman, 1998, page 98).

In essence there are two common features spanning all of the definitions. Firstly social capital is a concept external to individuals. It relates to and is applied at the society level. It is germane to the neighbourhood and society level, meaning it is an ecological construct. Secondly, a concept that bridges all the definitions of social capital is that they all incorporate a sense of public good (Berkman and Kawahci, 2000). It is a resource of rational agents needing to coordinate for mutual benefit (Woolcock, 1998).

2.4.6 Levels, types and forms of social capital

There are three different types; bonding, bridging and linking (Jochum et al., 2005). Bonding social capital incorporates the relationships between individuals that reinforce exclusive identities of homogenous groups (Harper, 2001, Khan and Muir, 2006). Putnam (2000) gave one example of bonding social capital whereby culturally similar people for exclusive organisations that inhibit integration with what are considered 'outsiders'. This type of capital is supportive to its members but often exclusionary to outsiders. Bridging social capital encompasses weaker inclusive ties that are associated with more distant relationships and networks that bring people together (Putnam, 2000).

The third type of social capital is linking capital. Linking capital corresponds to the power relations between groups and individuals in different positions of authority or society (Harper, 2001; Jochum et al., 2005; Khan and Muir, 2006). This can be extended to include the ability to leverage resources beyond the immediate community (Woolcock, 1998). Bonding and bridging social capital are applied to horizontal (peer) relations whereas linking capital introduces vertical relations between larger groups and organisations, as summarised in Table 3. "Bonding social capital constitutes a kind of sociological superglue, whereas bridging social capital provides a sociological WD-40" (Putnam 2000 page 19). Bridging and linking types of social capital are particularly relevant to debates around participation and governments because they can be associated to the different types of communication (Jochum et al., 2005).

Table 3: The three types of social capital

(Jochum et al., 2005)

It is these different forms of social capital that are often charged with positively contributing to society. It is believed to be a fundamental variable for various health outcomes (Subramanian et al., 2003; Bolina et al., 2003). It affects life expectancy and mortality (Kawachi and Kennedy, 1997), teen birth rate (Gold et al., 2002), and mental health or suicide (Araya et al., 2005). It has also been linked to crime (Kawachi and Kennedy, 1999; Sampson et al., 1997), economic development (Woolcock, 1998) and urban regeneration (Hibbitt et al., 2001).

2.4.7 Criticisms of social capital

Despite the large body of evidence and literature that supports the notion of social capital as a positive concept, there are many criticisms of the idea. The criticisms range from the lack of consensus on the theoretical definitions and terminology (Portes, 1998; Woolcock, 1998; Lochner et al., 1999), to the lack of standard measures (Lochner et al., 1999; Gaag and Snijders, 2004) which lead to different results from empirical analysis of social capital. Hawe and Sheill (2000) proposed the case that social capital is just another form of social support, and suggested that efforts made specifically to increase social capital by health promoters were redundant in light of other standard practices of promoting problem-solving capabilities. Perhaps social capital is a by-product of social action.

Alongside the more theoretical critiques some researchers have identified the negative effects of social capital from a practical view in society. Social capital, particularly in its bonding form can be exclusionary and lead to community isolation, social exclusion and the negative benefits of strong norms and reciprocity (Bauder, 2002; Baum, 1999; Portes, 1998).

The research is biased towards the positive consequences of social capital within society. There is very little emphasis or quantitative analysis that formulates the notion that social capital can contribute negatively to society and have detrimental effects on health, crime and wellbeing as well as impacting economic development (Portes, 1998). The homogenous and exclusionary nature of the bonding form of social capital often results in extreme negative consequences and development of insular communities and groups that limits integration and demands conformity.

2.5 Chapter summary

If the population centric model of health determinants is considered once again (see section 2.3.2), one key domain was the influence of the social attributes of the community. It is therefore pertinent to appreciate that the social attributes which are important for the success of the behaviour change programmes that run through social marketing can also be the social facts Durkheim sought to define. Social attributes include understanding networks, cohesion, influence, support and social change, together with social behaviours and lifestyles. Evidence in the literature links informal social networks, social activities and participation in organisations with better health chances (Cattell, 2001) , whilst some researchers observe the negative influences of social capital and its damaging effects on health (Baum, 1999).

The theory and measurement of social capital is still evolving and frequently debated, but it is still the most successful export of contemporary Sociology. Social space is composed of social, economic and cultural capital interacting at different levels and with different volumes for different geographical spaces over different time periods. If research could facilitate the understanding of inequalities and health status/outcomes in social space it would provide important contexts to a complex problem.

This first chapter of the literature review was interested in the policy context which provides an important frame for guiding research with real world application. Any understanding of health inequalities in the policy environment needs to recognise the importance of inequalities and their associated gradients in the population, meaning it is not appropriate to simply look at the absolute differences between rich and poor, healthy and unhealthy. Analysis and its resulting policy must consider the spectrum of inequalities across all population sub-groups resultant population gradients. The review of exiting models of social determinants of health helped gain understanding of the central elements that are part of the causal path of poor health status in populations and the different models indicated the importance of both the individual and there associated group.

Within the policy review, social marketing was identified as a re-emerging framework for delivering positive behaviour change impacting to increase health status of population sub-groups. The consequence of this naturally ties in with the priorities of tackling health inequalities that were identified. The main principles of commercial marketing are still applicable to socially responsible marketing which has been pioneered by social marketing practitioners. The process of which places emphasis on understanding the principles of product, price, place and promotion together with the 7 tenets of social marketing proposed by the NSMC to understand existing patterns of consumer behaviours. Inevitably these lend themselves particularly well to geographical analysis and segmentation to explore the most likely behavioural patterns for population sub-groups. This in turn leads organically to the requirement to identify differing health needs of differing population sub-groups. The forthcoming chapter considers these findings in more detail and turns to exploring how existing measures of health need are developed and used. To this end the next chapter includes a comprehensive review of geodemographic classifications.

Chapter 3

Literature Review

Part 2: Measuring health needs and expectations

3 A review of existing measures of health needs

The first chapter of this literature review identified the policy context within which this research fits. It then identified two theoretical backgrounds based on the types of models of social determinants of health, both of which are motivators for the existing policy. It then looked at the practical application areas into which this research will feed; health inequalities and social marketing. This next section explores existing research into how health inequalities, through the measurement of health outcomes/status, can be identified and targeted. This chapter starts with a brief appraisal of commonly used spatial and aspatial measurement techniques, and then considers in more detail the principles behind geodemographics and how they are currently used for measurement in the public sector.

Despite enormous improvements in technology and advances in material production, the effectiveness of public services delivery such as education, health and housing has not improved at the same rate as changes in material standards of living. As identified in the Wanless report (2002), standards of health service delivery have fallen behind other countries and the rise in morbidity and mortality related to diseases of comfort places even more pressure upon strained resources. Due to the recent decline in health service delivery, the presence of health inequalities and inequitable levels of local health outcomes, it is becoming increasingly important to develop robust indicators of health outcomes. If population groups in need of targeted health initiatives are identified and the level of need assessed, health professionals will be best placed to tailor their services to ensure they are equitable, efficient and, most importantly, successful.

Analysing and identifying population sub-groups with the largest disparities in health outcomes has become a priority for social marketers and health professionals alike. Identifying and reducing adverse health outcomes caused by diseases of comfort and health-harming behaviours is an important element of current policy and is important for reducing local and overall national disparities of health outcomes, by corollary the prevention of diseases of comfort has the potential to contribute to the freeing up of vital and rationed services and consequently help improve overall service delivery. But the complexity associated with identifying these types of health needs is evident in the literature. It is difficult to anticipate variation in characteristics of social-economic and lifestyle patterns of different population groups using existing health datasets. This is because many operational health datasets do not hold detailed information with any acceptable level of consistency or accuracy, and because of these measurements of health impacting behaviours related to lifestyle choices are rarely found at local levels of spatial aggregation. One of the paradoxes of local health care needs assessment is that data collected by general practitioners are not made available to PCTs for secondary analysis, because of the statutes of the 1998 Data Protection Act (OPSI, 1998) and laws relating to medical confidentiality (GMC, 2004) which protect data relating to the individual. The repercussions of this have led to a rise in the number of techniques and secondary surrogate datasets used to measure patterns and synthetic patterns in health inequalities and health needs some of which are discussed next.

3.1 Composite summary measures - area based deprivation measures

There are a number of existing estimation methods which predict prevalence of health related behaviours for small areas or that use proxy indicators for health outcomes and inequalities. One such proxy indicator considers the relationship between health inequality and deprivation through the creation of composite measures of deprivation to predict neighbourhoods of health need based upon levels of deprivation.

In both the models of social determinants of health presented in 2.3.2, it was evident that many compositional (individual) and contextual (group) factors interplay, giving rise to poor health outcomes. Such contextual factors include income and its close association with deprivation. But we must remember that deprivation is a controversial and elusive concept (Boyle et al., 2002), because its definition has been and still is contested. For the purpose of this literature review the notion of deprivation describes accessibility to material necessities for example adequate food and heating as well as social facilities affecting individuals, households, families, groups and institutions (McCulloch, 2001; Hirschfield et al., 1994).

It has become commonplace for health and other public sector professionals to use deprivation measures as proxy indicators for locating geographical areas with predicted poor health outcomes amongst the residential population. So what are they? Deprivation indices measure the proportion of households in a defined small geographical unit with a combination of circumstances indicating low living standards or a high need for services or both (Bartley and Blane, 1994). They are used to identify areas of relative concentration of material disadvantage, particularly in the absence of personal data (Slogett and Joshi, 1994). They reflect different aspects of living standards, personal, physical and mental conditions, local and environmental facilities, social activities and customs.

Throughout the 1980s and 1990s deprivation measures were in abundance and their construction and quantification became largely a statistical exercise. Most of the measures developed at that time utilised census attributes aggregated to small spatial areas. Variable distributions over spatial units needed to be consistent with preconditions for applying multivariate techniques so they were often transformed (ensuring variables approach a Normal distribution) and standardised (variables are scaled to make certain different variables have the same range) (Martin, 1996). The table below appraises some of the significant measures used throughout the 1980s and 1990s and the most commonly used measure at the time of writing.

Measure	Purpose of Indicator	Method	Limitations
Townsend Measure	Conceptualises poverty as a series of deprivations both material and social, where resources are so seriously below those of the average individual/household that the poor are in effect excluded from ordinary living patterns, customs and activities (McCulloch, 2001).	It is an un-weighted transformed combination of four census variables.	Townsend index has an urban bias and input variables are now too dated and not reflective of deprivation.
Jarman Index of Underprivileged Areas (UPA Score)	Designed to consider geographic variations in demand for Primary Care Services.	It was based on a survey of 100 London GPs and their subjective comments on social factors among the patients who most affect their workload. It uses 8 census variables at ward level. It is standardised against an average.	Not originally constructed to measure deprivation. Better at defining inner city deprivation because it includes factors like overcrowding and ethnicity (PAT 18). Bias towards London (Talbot, 1991). Two of the highest rated factors: proportion of population over 65 and transport difficulties, were dropped because they were already accounted for in the existing GP remuneration scheme.
Carstairs deprivation measure	Developed to assess Scottish health data analysis.	It uses 4 variables. Three are the same as the Townsend index, the forth variable looks at heads of households in a particular social class: IV or V.	Lack of theory surrounding choice of variables.
Breadline Britain	Consensual measures of poverty, based on what people themselves understand and experience as the minimum acceptable standard of living (Galobardes et al., 2006).	The index predicted poor households by using variables from the survey that are also present in the 1991 census. It is a mixture of data sources with variable provenance at the local level.	There is a high degree of inter-correlation between the alternative measures. This implies that the differences in design are small in comparison to the intrinsic differences of disadvantage and health care they are measuring (Field, 2000). Furthermore, associations between national surveys and census variables are imputed at the local level
The IMD2000/ 2004/2007	The ID2004 /2007 (and ID2000) were based on a multiple indicator, domain structured, approach to the measurement of area based deprivation, providing a basis for policy making (DCLG, 2004, 2007).	Makes use of data on persons or households in receipt of various government benefits made up of 6 domains (income, health and disability, education, skills and training, housing and access to services).	Undetermined relevance of variables in each domain to phenomena being measured. Each domain is given a weighting – unclear evidence on why weightings chosen. Arbitrary rankings of areas ranked in the inter-quartile ranges, therefore can only discriminate between the most and least deprived geographical areas.

Table 4: Table of commonly used measures of deprivation in UK

It is apparent that different conceptualisations of deprivation lead to different geographical representations of its physical and social conditions and as observed by Harris and Longley (2002) the way in which it is conceived determines the way in which it is measured. Table 4 outlines the application and limitations of each of the 5 measures reviewed. Each of the indicators used a number of variable distributions aggregated to arbitrary census boundaries to apply a ranked index score for the unit of analysis. For the Index of Multiple Deprivation 2004 (IMD 2004) this is the Super Output Area (SOA) of which there are 32,482 in England (a more detailed explanation of census geography is documented in Chapter 4).

If the distribution of data are analysed by area, as observed for both the deprivation indicators and the standardised mortality ratios discussed below, results will be influenced by choice of areal unit the analysis is performed upon. The choice of areal unit is critical due to its ability to be modified (Gatrell, 2001). As soon as areal units are used in reporting data and analysis the user faces the areal unit problem. Aggregation effects are associated with uncertainty about the scale required to represent a phenomenon (Martin et al, 2002). A more detail discussion of the areal unit problem is covered in section 3.5.4.

The areal units have considerable influence on the message. Indicators provided for large areal units at ward level do not always provide adequate detail. Thus, it is necessary to look at the smallest unit available for analysis. In terms of analysing Census data this equates to the Enumeration District (ED)/ Output Area (OA) which is never smaller than 16 households and 50 people, a general figure for urban areas. Even in this instance, the units of analysis often act only as containers of deprivation (Harris and Longley, 2002) and do not represent the complexity of compositional and contextual influences on health outcomes/status.

Another limiting factor of deprivation indicators is associated with their ranking. They are designed to identify neighbourhoods experiencing the greatest levels of deprivation, so they are not effective at differentiating deprivation levels of middle ranking neighbourhoods. Also some health-harming behaviours or diseases are not strongly correlated with deprivation or income, as such conventional deprivation measures (such as those listed in Table 4) would not useful for exploring such diseases of comfort.

3.2 Standardised ratios

One of the most widely adopted techniques in the UK for predicting generic health needs is to create standardised ratios of service use or morbidity for different administrative units, according to age, sex or level of local deprivation. This type of measurement is commonly found in health analysis carried out by both the Department of Health and the Office of National Statistics.

Limitations arise from using age-sex standardised ratios of service use, because they fail to address three key factors that are critical to health service planning: need, demand and usage. Simple age-sex standardised ratios of use do not account for actual health needs or demand because of the imperfect correspondence between service usage and health need. Variability in health outcomes can be ascribed to two main variable categories: composition and contextual. These are frequently used in the context of multi-level modelling where the variables correspond to units of analysis.

Compositional variables refer to the individual or the household for example the age and gender of the household whilst contextual variables correspond to differences arising from the group level attributes, for example the mean neighbourhood income of housing, which should not be reduced to an individual, for example absence of facilities in a neighbourhood. Age-sex standardised ratios reflect simple aggregations of isolated actions and do not incorporate more detailed compositional or contextual variables that influence health outcomes. What is more, Congdon (2006) suggests that because they are created using hospital data, it is likely that they confound health need with health supply.

The limitations of these age-sex standardised ratios were noted by Morgan et al. (1987) and were subsequently reiterated by Congdon (2006). Furthermore, aggregation of hospital utilisation data by administrative units may reflect different local coding practices at source hospitals or differences in referral rates and local policy mechanisms (Webber, 2004). Furthermore, if aggregated into small area extents, small sample numbers may invalidate the statistics. This thesis will explore whether geodemographic indicators fare any better.

3.3 Statistical and spatial statistical measures that can be applied to health need and outcomes

Other methods of predicting health outcomes and needs at the small area level may necessitate usage of complex spatially-enabled statistical techniques such as hierarchical or geographically weighted regression models to manipulate data and present geographical variations in health outcomes. Currently the Office of National Statistics collates statistics on a number of health needs – such as obesity, alcohol consumption and smoking prevalence – each of which is measured using the Health Survey for England (HSE). These synthetic estimates of healthy lifestyle behaviours were produced by the National Centre of Social Research (NATCEN). The synthetic estimates combined individual-level data from the survey with 2001 Census data. They calculate expected prevalence based on population characteristics using a number of ward level covariates as predictor variables within a multi-level logistic regression model (Pickett et al, 2005). As we have already seen, census wards are representative of an average of approximately 10,000 people, a unit of analysis that certainly does not accommodate local neighbourhood heterogeneity.

From a statistical viewpoint, a variety of regression techniques are most commonly applied to ascertain predictions of health needs. This is particularly common in epidemiology research. Complex hierarchical regression (multi-level) models and logistic models have been used for predicting health needs by studying individual epidemiological factors by incorporating simultaneously different levels of variables (e.g. family, neighbourhood, community) that influence the state of health (Rouz, 2003). Indeed there is a comprehensive research field that explores the application of multi-level models for epidemiology and modelling disease. It is beyond the scope of this thesis to conduct a comprehensive review of such models, but a systematic review of 25 studies can be found in a paper by Pickett and Pearl (2001) and an introduction to such models is provided in Goldstein (2003).

One such example of applying a multi-level model to explain variation in hospital utilisation because of health need in England was developed by Congdon in 2006. He explored variations in multiple outcomes by considering their totals for populations by area and GP practice using a hierarchical modelling approach. The same author also applied logistic regression techniques to estimate prevalence of mental illness using population sub-groups identified in national survey data. The predicted risks were then applied to disaggregated populations for the spatial areas of interest (Congdon, 2006).

In 2000 Twigg and colleagues analysed data derived from the Health Survey for England (HSE) and the 1991 Census in a multi-level model to predict health outcomes for fine scales. Using three levels of hierarchy, they attempted to replicate the individual behaviour of smoking and drinking by nesting individuals first within postcode sectors and then within local health authorities to produce ward-level estimates. They argued that "the simultaneously individual and contextual nature of the influences on health-related behaviour, provides the basis for the development of a new approach to the generation of small-area data on health-related behaviours" (Twigg et al., 2000, page 1111). The results of the multi-level model were then used to derive probability estimates of health-related behaviour for census wards, by predicting the proportion of people in an age-sex-marital status group who smoke.

Multi-level regression models (hierarchical / Bayesian models) can be a useful technique for ascertaining predictions at varying levels of a hierarchical structure such as patient- general practice – primary care trust, but they are not without their limitations. Often sample data contain small numbers which will skew the results, and often data are not available at the local level so predictions are made using aggregate data (the ecological fallacy).

In 2004, Asthana and colleagues also used the HSE to develop measures of relative class effect (RCE), for self-reported health. They used an adaptation of the location quotient concept (see below) to develop a measure to assess the relative roles of age, sex, and social class as factors underpinning these gradients. They called this the "relative class effect" (RCE). The RCE was defined as the, "ratio of the amount that class specific prevalence rates would have to change to leave all social classes with an equal prevalence rate, compared with the extent to which age specific prevalence rates would have to change to leave all age groups with an equal prevalence rate" (Asthana et al., 2004 page 304). This highlighted the importance of acknowledging that inequalities in morbidity are not homogenous across different social classes and gender. This research was conducted at the national scale, so caution must be applied to making assumptions at the regional or local scale, again given the possible emergence of ecological fallacy.

In economic geography a location quotient is used to compare the local economy to the larger regional or national economy. It is a ratio expressing the industry's share of the local economy compared to the industry's share of the regional/national economy. They are used to determine the concentration of a particular type of industry.

In health needs analysis location quotient techniques have been adapted to produce a neighbourhood quotient model of health needs. This particular technique ascribed to health outcome measurement can be attributed to Webber (2004) and Webber and Longley (2003), who produce neighbourhood quotients of expected patterns of health needs based on different diagnoses produced by hospitalisation data. Webber (2004) took data from the national Hospital Episode Statistics (HES). By joining patient postcodes to a national geodemographic classification he produced national demographically standardised ratios of expected hospital usage. He created ratios of the neighbourhood patterns of hospital usage per diagnosis which were compared to the England average per diagnosis. The ratios were stratified according to the geodemographic Type. They differ from traditional location quotients because they do not aggregate data into coarse geographic aggregations but rather gather strength through social similarities as defined using a geodemographic classification.

Most recently spatial micro-simulation has been used to predict urban and regional populations and their variation in health outcomes. Micro-simulation is concerned with the creation of large-scale datasets for the purpose of estimating attributes of individuals within households. Ballas and colleagues (2006) have used multiple years of the population Census combined with the British Household Panel survey to create a range of conditional probabilities which are subsequently used to simulate population projections up to 2021, at the census ward units of analysis, which are likely to be too geographically extensive for social marketers to be specific enough.

More geographically centric methods are also evident in the literature which include the following methods: Geographically Weighted Regression (GWR), cluster detection, kernel density smoothing, spatial autocorrelation (Moran's I, Geary's C), Local Indicators of Spatial Association (LISA), spatial regression and even spatial filtering (a process commonly used in remote sensing) some of which will be discussed below in more detail. A review of some of these techniques and their application to public health data has been conducted by Chung et al. 2004.

Some of the techniques identified above are concerned with exploring the presence of spatial autocorrelation in a dataset: LISA, Moran's I and Geary's C. Spatial autocorrelation is a term used to consider data from locations spatially near to one another which are more likely to be similar than data from locations remote from one another (O'Sullivan and Unwin, 2002, page 28). In other words near things are more similar than far things (and are therefore spatially correlated), known as the first law of geography (Tobler, 1970), "Everything is related to everything else, but near things are more related than distant things) (page 236).

The definition of spatial autocorrelation is, "The degree of relationship that exists between two or more (spatial) variables, such that when one changes, the other(s) also change. This change can either be in the same direction, which is a positive autocorrelation, or in the opposite direction, which is a negative autocorrelation", (AGI, 2007, online) The exploration of spatial autocorrelation in a dataset looks for non-independent observations in the spatial domain and can be assessed using a number of geo-statistical measures which determine either global or local spatial autocorrelation. These indicators of spatial autocorrelation provide one mechanism for exploring Tobler's First Law of Geography to determine if data patterns really do exist.

Global measurement of spatial association is typically explored by using either Moran's I or of Geary's C (Fotheringham et al, 2002) statistics. The Moran's I statistic is a weighted correlation coefficient often likened to the Pearson's Product Moment Correlation Coefficient used in conventional a-spatial statistics. Both statistics identify non-random data distributions at the global scale by comparing the statistic returned to expected values for the random process. A comprehensive explanation of both statistics and their applications is detailed in O'Sullivan and Unwin, 2002, chapter 7). Their application at the global scale only indicates whether or not the overall data distribution is spatially auto correlated but what they do not explain is any significance in local variations, because of the specific local variation in health outcomes, it is critical to identify local data anomalies.

Local measures of spatial association are concerned with exploring the presence of geographical clustering at the local unit of analysis. Local spatial autocorrelation can be investigated using two common methods: local indictors of spatial association (LISA) statistics or geographically weighted regression (GWR). The LISA statistic was developed by Anselin (1995), and is the local equivalent of Moran's global I statistic. It explores the similarity of an area compared to its neighbouring areas, and then tests the strength of this relationship. As described by O'Sullivan and Unwin (2002, page 203), "LISA statistics are disaggregated measures of autocorrelation that describe the extent to which particular areal units are similar to or different from, their neighbours". They are exploratory statistics that enable the researcher to determine if an area differs unusually from its neighbour. A review of spatial statistics for health can be explored by reading Waller and Gotway (2004).

Geographically weighted regression enables local as opposed to global models of relationships to be measured and mapped. It does this by incorporating the local spatial relationships into a standard regression framework. It was a geographical statistical techniques developed in the 1990s by Brunsdon, Fotheringham and Charlton (1998). Examples of GWR applied to the health domain and predicting health outcomes have been explored in a growing number of journal articles, but were first considered by the creators of the technique in an article published in the Statistician (1998). In this article Brunsdon and colleagues applied the newly conceptualised GWR process to predict prevalence of long term limiting illness based on a number of compositional and contextual variables taken from the 1991 Census.

This section provided a basis for identifying the common spatial and aspatial techniques used to explore and predict patterns of health outcomes, each with their own benefits and limitations. The most significant with respect to social marketing application is the scale at which they are made available, is often too large to identify the needs of specific homogenous population sub-groups. Confounding this is the statistically complex nature of some of the methods GWR and (spatial) multi-level models ensure the techniques remain in academia and within a small group of experts. In order to create effective and useful metrics that are accessible to a wider audience then simple but robust methods need to be investigated. Geodemographics as small area measures of socio-economic and lifestyle conditions may prove themselves useful for this task. The applications do not set out to find explanations for patterns, they do not attempt to consider causality, but they do provide population differences to be used for exploratory purposes and identify areas of further investigation. The research agenda of this thesis investigates the mechanisms through which geodemographics can provide a vehicle for exploring distributions in health outcomes related to lifestyle behaviour. For this reason this next section reviews the current research and thinking relating to geodemographics; what they are, how they have been used and their strengths and weaknesses.

3.4 Geodemographics

Regardless of the model used to explain health inequalities and or deprivation there appears to be an inherent hierarchical organisation and spatial dimension to their determinant variables or factors; with variables interplaying at the level of the individual and the group: i.e. social patterning. The review of social marketing in the previous chapter noted the inherent geographical component within the traditional principles of marketing: product, price, place and promotion. This together with 7 principles that comprise the UK framework for social marketing: insight, exchange, competition, intervention, segmentation and behaviour. It seems that geodemographics used correctly with the most relevant data could be a useful tool for providing information and knowledge for a number of them.

The previous section identified an array of measures used to describe and/or predict the nature and pattern of health inequalities, deprivation or health outcomes. The techniques used changed in accordance to the availability of data at the time, the purpose for the analysis and the expertise of the user. The next section discusses an alternative method, commonly used in marketing and finance, to identify different neighbourhoods and types of people: geodemographics.

3.5 What are geodemographics?

Studies of demography investigate both the individual and group characteristics to explore behaviour and lifestyle patterns, which fits in with the desire of the researcher to explore variations in diseases of comfort and other preventable diseases. The underlying principle upon which geodemographics are constructed considers that similar people live in similar types of neighbourhoods, go to similar places, do similar things and behave in a similar manner as in the old adage; "Birds of a feather flock together".

The etymology of the word geodemographics come from two Greek words: geo (earth) and demos (people). The first geo is an abbreviated version of the word geography that is concerned with the arrangement of places and physical features of the Earth. The second word, demography, references the study of populations - placing particular emphasis on size, density, structure and characteristics. It is the study of people in relation to place, investigating the combination of demographic data and information with relation to their linkages in space. It is a term applied to the analysis of social and economic data in a geographical context (Johnston et al, 2000) and is commonly referred to as the analysis of people according to where they live (Sleight, 2004). It suggests that by knowing where someone lives, it is possible to say something about the characteristics of that person or group of people.

Geodemographic typologies are a popular method for identifying different population groups used in the commercial sector; national demographic databases are routinely used in finance and marketing industries to classify both lifestyles and behaviours geographically. These typologies categorise individuals into different types and groups of people according to similarities in their socio-economic circumstances, lifestyles and behavioural patterns. They act as a segmentation tool for customers/patients. This enables more efficient, effective and exact marketing initiatives to be developed. In the private sector their use is not just restricted to marketing, they have been widely applied to assist survey design, retail planning as well as direct marketing (Harris et al., 2005). Using these typologies, the fields of finance and marketing are better placed to understand the profitability of different customers, the ideal location for new premises, the potential demand for different products according to different locations and the people who should be targeted and correspondingly where these people live.

Longley et al. (2001) identified that geodemographics were and indeed are habitually used in business to provide operational, tactical and strategic context to decisions that involve the fundamental question, where? This allows marketers to build knowledge about the place aspect of their campaigns. Recently the public sector has become interested in applying geodemographic typologies to enhance understanding of population need, demand and use of different services. Of late there has been a renaissance with respect to the application of geodemographics within the public sectors, which will be discussed in detail in section 3.5.3. There is a lot of potential for them to lend themselves quite naturally to social marketing.

The geographical context of 'Birds of a feather flock together," is provided by Tobler's First Law of Geography, which was first discussed in section 3. The law states that everything is related to everything else, but near things are more related than far things; the technical term is spatial autocorrelation (SA). The previous section noted that spatial autocorrelation quantifies the degree to which near and more distant things are interrelated, a phenomena can be classified as either having positive or negative spatial autocorrelation. Positive SA means similar values will appear together. As identified by (Harris et al., 2005), geodemographics inherently exhibit positive SA because residents in the same neighbourhood are more likely to possess similar socio-economic and lifestyle characteristics.

3.5.1 Founding theory of geodemographics

Both Harris et al.(2005) and Batey and Brown(1995) have detailed the development of geodemographics from its theoretical beginnings; tracing it back to the Chicago school of urban sociologists in the 1920s. Much earlier in 19th Century London, the philanthropist Charles Booth set out to produce a street-level poverty map of London, because he was unhappy with the existing quality of poverty measurement. He conducted a survey of life and labour in London during 1886-1903, using qualitative techniques such as focused interviews and observations (Figure 7).

Figure 7: Booth poverty map

(LSE, 2006, online)

During this survey he interviewed many different types of Londoners and maintained fastidious notes describing the city street by street (LSE, 2006, online), as shown in Figure 7. He was able to develop a standard classification of London's streets based on his survey which classified the streets according to social class and income.

Moving slightly forward in time, during the 1920s urban sociologists and geographers were concerned with investigating the establishment of general principles about the internal spatial and social structure of cities (Batey and Brown, 1995). There were three principal sociologists; Burgess, Parks and McKenzie who set about investigating human behaviour in the urban environment. In 1925 Burgess set about dividing the city into a series of concentric rings, with each ring designated as a different land use zone which attracts different types of social groups. In Figure 8 the central zone is referred to as the central business district (CBD) where all the major transport routes emanate.

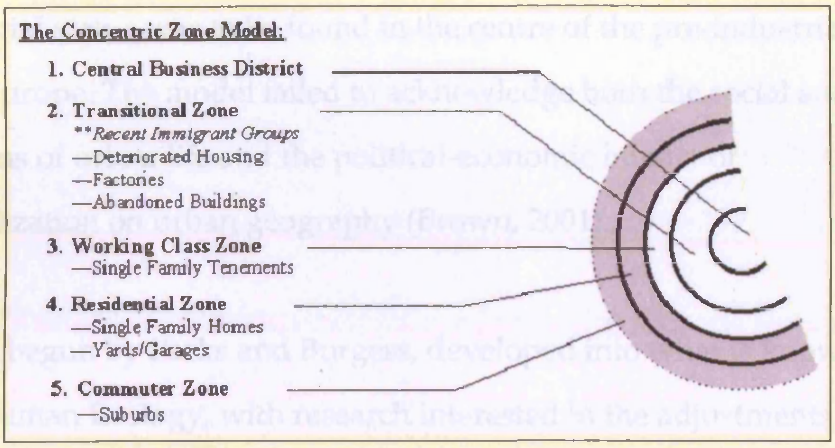


Figure 8: Burgess urban land use model

Data source: (Crime Theory, 1998, online)

These urban sociologists defined natural areas within cities; outcomes of an unplanned spatial sorting process of similar people through the operations of the housing market (Johnston et al., 2000). The natural areas are highlighted in Figure 8. Burgess and Parks then explored the relationships between urban characteristics and the social, economic and cultural characteristics of the population. This enabled exploration of the relationship between social control and competition.

The theory used for the urban land use model was based on the premise that as people struggled for scarce urban resources such as land, competition grew between different groups. This lead to the division of urban space into ecological niches or natural areas where people shared similar social characteristics because they were subject to the same ecological pressures (Brown, 2001).

The model of urban land use, though simplistic in quality, was not without its flaws, it assumed spatial separation of the work place and the home. It was also developed using information about American cities; its application to European cities was not always appropriate because affluent housing and higher social status was to be found in the centre of the pre-industrial urban areas in Europe. The model failed to acknowledge both the social and cultural dimensions of urban life and the political-economic impact of industrialization on urban geography (Brown, 2001).

The work begun by Parks and Burgess, developed into what is known as the field of Human Ecology, with research interested in the adjustments made by man to his physical environment. Human Ecologists believe that people must be studied in their own environment. Batey and Brown (1995) noted that the growth of Human Ecology was stimulated by the increasing availability of data, particularly census tract data, when detailed population data become accessible.

During the 1950s Shevky and Bell in the United States developed a technique known as social area analysis, making use of the available census data. The work was considerably influenced by the Chicago School of Human Ecology. The key aim of their work was to relate measures of social change to the geographical structure of the city through the identification and description of areas based on their social characteristics. Thus, the technique of social area analysis linked urban social structure and residential patterns to economic development and urbanisation processes. The theory considered both social change and the definition of a complex social scale, enabling the identification of relatively homogenous areas with respect to housing conditions and the social environment, linking the changing urban society to residential differentiation within urban areas.

Early development of social area analysis had a slow start in the UK and it was stalled by the lack of availability of small scale census data, but this changed with the 1961 census. Following the release of the 1961 small scale census data, in 1969 a classification of enumeration districts was produced as part of the Third Survey of London Life. Enumeration districts were classified into one of six different types; upper class, bed sitter, poor, stable working class, local housing authority and almost suburban (see Figure 9). It produced one of the earliest examples of geodemographics.

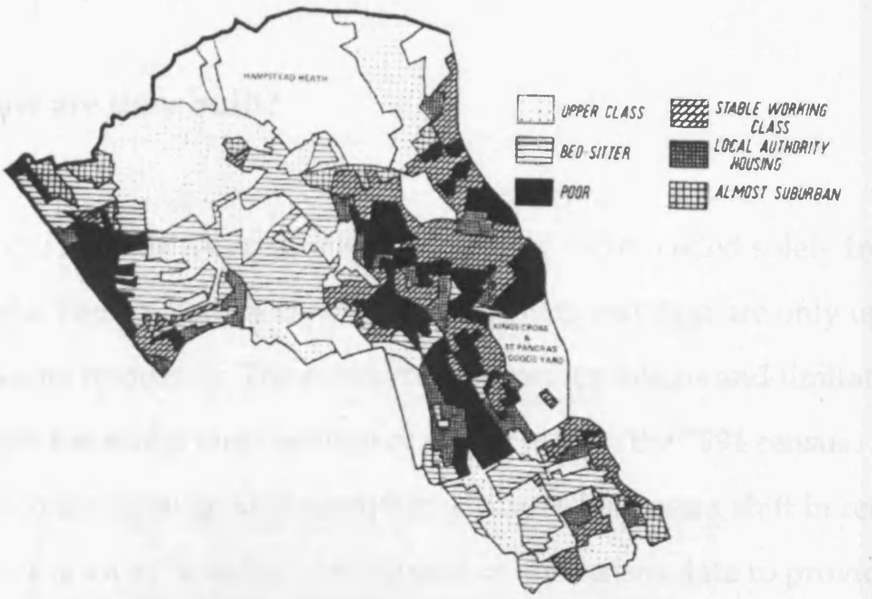


Figure 9: Geodemographic classification of Camden from the 1961 census

Data Source: (Doogan and Rokkan, 1969)

This early crude geodemographic classification system illustrated the potential usefulness of using attribute similarity to develop classifications of population groups to explore variation. At the same time the Liverpool Social Malaise Study was being carried out. The study sought to identify areas within the city where there was a high incidence of social malaise, by using social statistics to combine census data and operational data held by Liverpool Council.

This work was the precursor to a national classification system that was developed by Webber whilst at the Centre for Environmental Studies. Wards, parishes and local authorities were classified according to national averages. The Parish and ward classification typology became ACORN, the first UK neighbourhood typology. From this point on geodemographic typologies in their current form were born in the UK, due to the availability of small area data and increased computer processing power. To enable understanding of how a geodemographic typology works, it is relevant to briefly describe the processes that are used to build a geodemographic typology.

3.5.2 How are they built?

Geodemographics as we are familiar with today were seeded solely from census data. The UK census is held every 10 years and data are only updated with the same frequency. The census has its own problems and limitations, for example the under enumeration of young men in the 1991 census. As a result the commercial geodemographics industry has seen a shift in recent years, moving away from its total reliance on the census data to provide information to build typologies. They have moved towards the integration of other commercially available types of socio-economic and lifestyle datasets. Through the augmentation of census data with other datasets they facilitate the process of improved understanding of the social geography of small areas and add value to the underlying data (Brown et al., 2000). Examples of the datasets that add value include: County Court judgements, lists of company shareholders, council tax bands, the British Crime Survey and the electoral register for example.

Despite this move to incorporate other datasets by commercial providers, the UK Office of National Statistics has developed in conjunction with Leeds University a geodemographic classification system of output areas (SASIRG, 2005). The classification used 41 transformed and standardised census variables to classify every output area in the UK based on its value for those variables. A detailed methodology can be obtained from the Office of National Statistics (ONS, 2005), which has an advantage over the commercial providers because of its open and transparent methodology which can be examined and critiqued.

Commercial geodemographic segmentation systems take millions of raw data records. They divide households into groups and types based on similarities in income, education and household type together with information about lifestyles, attitudes and product preferences. They create a classification system for local neighbourhoods based on the postcode unit. These classifications are then used to ascribe characteristics to populations.

In the literature an extensive debate surrounds the meaning and definition of neighbourhood. It is beyond the scope of this review to discuss its meaning, but within this research the contextual meaning of neighbourhood refers to a small area that equates to the Royal Mail postcode unit. These units of analysis are discussed in more detail in Chapter 4, section 5.1.2.

Biological taxonomy is "the classification of organisms into a hierarchy of groupings, from the general to the particular, that reflect evolutionary and usually morphological relationships: kingdom, phylum, class, order, family, genus, species" (Encyclopaedia Britannica, 2007). In the same vein geodemographics repeat this practice with households, streets and neighbourhoods, with the view to providing greater understanding of the diversity of neighbourhood composition through the creation population taxonomies/typologies/classification (these terms are often used interchangeably).

In a similar way that a species is a "subdivision of biological classification composed of related organisms that share common characteristics and can interbreed" (Encyclopaedia Britannica, 2007), a geodemographic type is a cluster of households that on average share socio-economic characteristics that are dissimilar from other types. Geodemographic typologies are clusters formed by grouping areas that appear to be the most alike together. They can be applied despite the political or administrative boundaries of an area and are built using a more or less similar methodology (Dramowicz, 2004).

Harris and colleagues (2005) identify a seven stage process involved in the creation of geodemographic typologies. The first step requires the selection of potential input variables. The input variables must be relevant; in the case of commercial typologies census data are augmented with other available datasets, e.g. the electoral register, county court judgements etc., providing greater granularity of information. The second step brings data together in a common geography, permitting the evaluation of the data – whereby variables are transformed to a Normal distribution (step 3). Once transformation is complete, the variables are weighted (standardised) (step 4). Clustering of the variables begins at step 5. At this point it is necessary to select the number of clusters to be produced. Dramowicz (2004) suggests that the number of segments = 1+3.3 log (number of objects). In which case a geodemographic typology based on the 1.6 million postcodes in the UK should have approximately 25 segments. Although in the literature there appears to be no hard and fast rule as to the number of clusters.

The clustering of variables can be carried out using one of two ways; using a top-down method or by allocation-reallocation (K-Clustering). Experian use the allocation-reallocation method with a K-Clustering algorithm to produce their geodemographic typology, Mosaic. The OAC classification also uses K-means clustering to produce the Output Area Classification. Following completion of the clustering the final steps involve arranging the clusters into a hierarchy and assigning a numerical, textual and visual summary of the clusters. The objective of cluster analysis is to consign objects into groups or clusters inherent within the data structure. The groups are not previously defined in any way. Objects in a given cluster tend to be similar to each other in some sense, and objects in different clusters tend to be dissimilar. Cluster analysis can also be used for summarising data rather than for finding natural or real clusters. This use of clustering is sometimes called dissection (Everitt, 1980), and is the basis of geodemographics.

3.5.3 Applications of geodemographics

Following the birth of ACORN ('A Classification Of Residential Neighbourhoods), the commercial geodemographics industry exploded in the UK, with the emphasis placed on business, finance and marketing and more recently interest has grown in the public sector because of collaboration with universities and active marketing of the tools to public sector agencies. The two market leaders for these technologies in the UK are Mosaic and Acorn. Mosaic claims to be used by over 10,000 organisations world wide (Experian Business Strategies Division, 2005) whilst Acorn claims to be the leading geodemographic tool used to identify and understand the UK population and their demand for products and services (CACI, 2005).

Birkin and Clarke (1998) wrote about the use and potential use of geodemographics within the financial services industry following the growth of both the purchase and use of financial services by individuals. Within this industry there are many uses for this type of typological system. Geodemographic typologies assist the finding of new, and profiling of existing customers. Catchment area analysis is conducted to facilitate the decision-making process that determines the most effective branch locations. Although in the mid 1990s analysis was relatively crude, GIS was used to overlay population data onto areas with buffered catchments. More recently, use has been made of gravity models and geodemographics to assess location - allocation problems where consideration of the distance from a location, the attractiveness of a site and the competition in the surrounding area are all accounted for.

By comparison, the use of geodemographics in the public sector has been growing modestly since the mid 1990s. Growth primarily began in the public sector within education. Typologies were used to ascertain the characteristics of students accessing Higher Education (Tonks and Farr, 1995; Tonks, 1999) and nowadays they are being used to understand the socio-economic groups underrepresented at Higher Education institutions (Singleton, 2004) to allow decision-makers to understand the complexities involved with wider participation.

In 2005 Ashby and Longley identified the importance of small area geodemographic profiles, noting that they are pivotal to the tactical and strategic resource management in many areas of businesses. They are becoming central to the efficient and effective deployment of resources by the public service. They propose that the application of geodemographics within local police environments provides significant community intelligence. This work builds upon studies first presented by Brown (see Brown et al., 2000), who identified the application of geodemographics typologies within community safety to analyse and interpret patterns of criminal and anti-social behaviour, enhancing the evaluation of effectiveness of responses by different agencies to telephone calls reporting incidents.

Current interest in equity of services by both public sector bodies and researchers has led to a growth in research projects that use them to understand population differences. A recent study in the United States used geodemographics as a tool to stratify the population, to assess the geodemographic correlates of broadband access and availability (Grubesic, 2004).

The interest in utilising geodemographic typologies for health issues is not new, many research have explored the use of segmenting populations to understand postcode health databases (Brown et al., 1991; Hirschfield et al., 1995; Openshaw and Blake, 1995), suggesting that typologies provide a useful descriptive tool for the analysis of health data whilst outlining the limitations. More recently (Webber, 2004), investigated the neighbourhood inequalities linked to patterns of hospital admissions and how they could be used to effectively target health promotion activities. Each of these applications of geodemographic typologies whilst being beneficial also have their limitations, premises and caveats that must be applied and understood before using them.

3.5.4 Limitations of geodemographics

It is difficult to review comprehensively the techniques used to build commercial geodemographic typologies. Companies responsible for building these systems do not release sufficient technical detail because they wish to maintain their intellectual property or gain the competitive edge. As a result, it is problematic to critically appraise and assess their validity, robustness and quality.

There are three distinct types of limitations that are present in the geodemographics industry. Firstly uncertainties are introduced during the technical construction of the typology. Secondly external uncertainties are caused by statistical aggregation, and finally the third type of uncertainty occurs when geodemographic typologies are applied to other operational datasets.

As with all systems and analysis, geodemographic systems and typologies are only as good as the input data used to build them and error and uncertainty is introduced during the technical development stage of system. A geodemographic typology is built by clustering many variables based on their similarity. In commercial geodemographic typologies the actual list of diagnostic variables and the weightings given to them for input are not in the public domain. Consequently appraisal of the quality, completeness and accuracy of these variables is not possible and one must assume that they are fit for purpose. Running parallel to this concept is the problem posed by the decision behind choosing the input variables. It is not possible to identify the epistemology that underlies these decisions. Because of the open source nature of the Output Area Classification (OAC) classification and its non-commercial nature, the details of the methodology and input variable are in the public domain.

During the clustering process, the small areas that are grouped together may not be adjacent; by corollary the characteristics of an individual postcode unit are likely to be lost in the averaging process. In the current geodemographic typologies, a postcode unit can only belong to one group or type. This means that the probability of the postcode actually having different characteristics as opposed to the descriptive statistics of the cluster is quite likely. The effect is also known as the scale effect of the Modifiable Areal Unit Problem, discussed in more detail below. The clustering process itself brings with it its own complications, limitations introduced are dependent on the algorithm used to produce the clustering but a discussion of these limitations are outside the scope of this review as they are concerned with detailed technicalities of clustering methods.

Within the field of spatial analysis, two problems impact on aggregated statistics of all types: the ecological fallacy and the broader issue of the modifiable area unit problem. Georeferenced data are more often than not aggregated together to form different types of zones (units of analysis), for example, in the UK, government statistics are grouped together into units representative of different administrative boundaries: wards, police sectors, parliamentary constituencies or local boroughs. These analysis units are arbitrary.

This type of unit aggregation eliminates issues surrounding that of individual identification, and provides decision makers with simple understandable statistics, but creates the classical problem of ecological inference (Weeks, 2004), which results from aggregation bias. The ecological fallacy relates to the fact that observed group-level associations are often used to make inferences about individuals that are not necessarily applicable. This results in the drawing of incorrect conclusions from aggregated data. This issue was first identified by Gehlke and Biehl (1934) whilst examining census tract data. They discovered that whilst there was a strong relation between the race and literacy at the census tract level, detailed examination at the individual level illustrated that there was no relation between the two variables.

The modifiable areal unit problem (MAUP) is a broader complication than the ecological fallacy, which is just one facet of the MAUP, and is a serious issue within geodemographic research (Weeks, 2004; Debenham et al., 2003; Openshaw and Blake, 1995). The MAUP can be divided into two: the scale effect and the zoning effect. The scale effect comes into play when variation in results are obtained when data from one set of areal units are progressively aggregated into fewer and larger units of analysis (Openshaw, 1984). How this manifests itself was identified by Debenham (2003), who observed that during the creation of geodemographic typologies, clustering algorithms group small areas together for example output areas or postcode units.

The second facet of the MAUP relates to the zoning issues and the complications that arise through their aggregation. Quite simply it references the ecological fallacy discussed above, whereby the aggregation of different zones of the same data, will produce a multitude of results. The complications raised by the ecological fallacy and the multiple areal unit problem are not unique to the geodemographics industry and are faced daily by GIS experts in academia, the private sector and the public sector.

Curry (1997) argues that places are seen as locations to which individuals are contingently attached. "Lifestyle" profiles are nothing more than statistical aggregations. A notion exacerbated by the lack of internal validation of the classifications. He argues that they commonly miss the relationship between the individual and the place or neighbourhood, noting that the place is simply the location with a set of individuals attached to it, representing crude approximations. Goss (1995) pointed out that social behaviour could not be reduced to an aggregation of measurable demographic and physcographic characteristics and are classified into a limited number of types.

Within marketing, social network and social organisation theory there is a concept known as the homophily principle. It is an organising principle that suggests that,

> *"...contact between similar people occurs at a higher rate than among dissimilar people. The pervasive fact of homophily means that cultural, behavioral, genetic, or material information that flows through networks will tend to be localised. The result is that people's personal networks are homogeneous with regard to many sociodemographic, behavioural, and intrapersonal characteristic."* (McPherson et al., 2001, page 416).

Geodeomographic typologies set about identifying these networks of association and similarity of behaviour to enable effective communication. They facilitate the identification of homophilious networks. Longley and Webber (2003) suggest that there are location effects that arise out of geographical proximity which complement the effects of social similarity that occur in geodemographic clustering. So whilst they may be subject to a form of the ecological fallacy, because they are concerned with averaging, they may provide a useful framework for social marketers because they provide enriched data at the local scale.

3.5.5 Ethics of geodemographics

One further complication of geodemographics is that they are often difficult to quantity, and this relates specifically to the ethical implications of their use. The need for ethical consideration in geodemographics systems has been consistently identified since the mid 1990s. Curry, 1997 identified the proliferation of national databases by Curry and how they enable compilation of extensive dossiers on individuals. In the case of the UK examples include; the Crime Recording Information System that contains records about every recorded incident and classified crime in England and the Exeter patient registration recording system. Combining all this information with geodemographic typologies provides an incredibly detailed picture of the population. He suggests that this leads to a re-conceptualisation of the nature of the individual, relating it to the ecological model of urban science in which residential locations reveal primary social and cultural characteristics of individuals. Curry is not the only one concerned with the proliferation of these systems. In 2005 the Joseph Rowntree foundation (Burrows et al., 2005) expressed concern about the dissemination of this type of data over the internet. Calling such internet based systems as a 'social sorting' technology. They argue that, "despite policy to encourage 'mixed communities'– processes of social differentiation and fragmentation are intensifying ....such systems have the potential to change fundamentally or to solidify the image of individual neighbourhoods" (online). The ultimate consequence of such systems might reinforce social fragmentation.

Despite the limitations and uncertainties attached to the use of geodemographics, the typologies have stood the test of time, and have been used in the commercial world successfully for over 20 years. It would seem that commercial geodemographic systems provide the user with an easy to use opportunity to understand more about the location of different population types at low spatial resolutions.

Limitations are placed on industries to ensure the confidentiality of individuals within the constraints of the Data Protection Act 1998; geodemographic typologies offer a method of aggregating large numbers of variables maintaining sufficient richness and detail of population data whilst retaining anonymity. Geodemographic typologies provide a method for generalising populations – into units of analysis that provide sufficient detail for strategic decision making. What is apparent from the review of existing methods and potential methods for exploring existing patterns and predicting expected patterns, geographically appropriate estimates of health outcomes and behaviours that reflect social similarities of local populations are useful.

3.6 Summary of the literature review

Gradients of inequalities traverse the social spectrum and are in no way merely restricted to the poorest members of society. Health and environmental disparities are not exclusive to any one population group, community or geographical area. The 2004 Public Health White Paper produced by the Department of Health reiterated that the differences in health between neighbourhoods and socio-economic groups are unacceptable (Department of Health, 2004).

It is in this context that much recent research has set out to measure and to quantify health inequalities by linking it to material and income deprivation. The literature identified that local geographies of deprivation have been used to identify areas which lack basic to material necessities and access to facilities. A range of indicators have been used to measure 'deprivation', as a broad proxy for poor health status, for example: the Townsend Measure (Townsend et al., 1988), Jarman Index of underprivileged areas (Jarman, 1983 and 1984), Carstairs' Index (Carstairs and Morris, 1989 and 1991) Breadline Britain (Gordon and Pantazis, 1999), and more recently the Index of Multiple Deprivation (IMD) (DCLG, 2004, 2007). Despite this, it still remains a difficult concept to conceive, measure, and analyse because of its heterogeneity and the high degree of uncertainty around such representations (Longley and Harris, 2002).

The review noted that despite the difficulty in defining deprivation the associated measures have nevertheless been used in conjunction with a number of other techniques in order to assess geographic variation innate to health needs assessment and equity analysis. A range of measures have been used to specify and estimate both people effects (compositional) and/or contextual 'place effects'. To determine the extents to which a neighbourhood reproduces health inequalities. Previous statistical measurement methods used traditional standardised age-sex ratios, logistic regression or advanced multilevel models with a combination of census and or survey data (Twigg et al., 2000, Pickett and Pearl 2001, Asthana, 2004, Congdon, 2006,). More explicitly the geographical approaches that were reviewed manipulated spatial simulation modelling (Ballas et al., (2006), engaged in geographically weighted regression (Fotheringham, Charlton and Brunsdon, 1997), or other forms of spatial regression (Anselin, 1995, Waller and Gotway, 2004). Many of these methods required advanced statistical knowledge to compute and interpret, are technocratic in nature and difficult for the lay person to understand and interpret.

Each of the techniques for measuring inequalities, health outcomes, needs and equity highlighted societal disproportionalities. The measurement of which has become more evident with the advent of improved data and demographic information the effect of which led to the development of new techniques and processes for shaping, modifying and changing health related behaviour(s). New policy indicates the importance of improving behaviours in order that they might be beneficial to society, under the auspice of social marketing (section 2.4), which ultimately aims to the reduce inequalities.

The review noted that over the last 25 years, the profile of social marketing has risen slowly up the political agenda. The most recent Public Health White Paper (Department of Health, 2004) which placed significant weight on health in a consumer society and the requirements for local communities to take action to improve health.

From exploring the literature it was evident that empowerment of individuals and neighbourhood communities to take positive action for reducing local inequalities and changing health behaviours has at least two main prerequisites. Firstly, local participation and action in the health agenda must be stimulated through the provision of understandable, relevant and appropriate information about local inequalities to local communities. Secondly, a health promotion and prevention initiatives needs to be embedded into existing programmes and informed by relevant information about priorities.

A review of the current policy climate indicated that the UK public sector is under considerable pressure to improve standards via the provision of efficient, effective and equitable services that best meet needs and offer improved value for money. One solution proposed is to tailor services to be more specific to population groups and consequently reduce the prevalence of diseases of comfort. If the NHS and its partner organisations are successful, it is thought this will reduce the levels of existing health inequalities.

The population centric model of the health inequality determinants proposed by Friedman et al. (2002) placed importance on understanding social similarities across population sub-groups through the identification of community attributes, collective lifestyles and health practices. With an appreciation of social similarity across and corresponding health outcomes across population groups, health professionals are better placed to create germane interventions through application of social marketing. There is a definite research gap within social marketing and health inequalities fields for more easy to use, robust measures that could be filled by geodemographic applications.

A review of the existing measures of health comes, identified a lack of simple small area measures which are able to predict differences in health outcomes/status. Although there is an abundance of complicated statistical models, which are impractical for every-day use in the health care setting by health care professionals. This is because they require specialist skills and knowledge. On the other hand geodemographic modelling of health outcomes provides an alternative approach. They offer a spatially disaggregated data representing small area levels of social similarity, and are potentially useful for regional and social policy analysis.

Chapter 4

Research framework, aims and objectives

4 Research framework: aims and objectives

4.1 Literature summary

The literature review synthesised reading from three key areas: inequalities, social marketing and geodemographics and included a brief synopsis of social capital, as illustrated in Figure 10. It provided a synthesis of the contextual underpinnings of the research through the health inequalities literature and identifies a primary application domain for the research outcomes in the field of social marketing and social capital. It was evident that the current policy climate requires a focus on evidence –based decision making and appropriate services and health improvement/preventative initiatives targeted at population sub-groups. The research highlighted the importance of exploring health outcomes for different segments of the population.

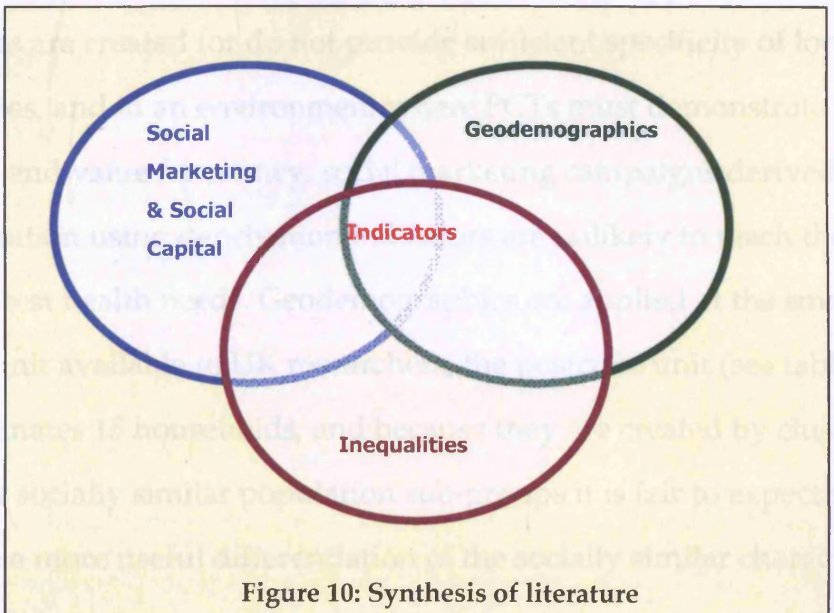


Figure 10: Synthesis of literature

Inequalities in the health domains are inherently spatial because the individuals or communities are located in particular places. Health outcomes vary from person to person and from area to area. They can even be a result of the place in which one lives (for example the temperature of area can cause illness). For these reasons they are inherently geographical in nature (Marmot and Brunner, 2005). The literature review highlighted the extent of inequalities in the UK and outlined some of the many goals related to the reduction of inequalities. The variation in conditions often relates to disparities between particular social groups which are associated for example with gender, age, income, lifestyle, geographic location and ethnicity.

4.2 Problem definition

The systematic literature of the three topics: demographics, social marketing and health inequalities identified a number of gaps in the research. There is a requirement for health outcomes, needs and inequality measurement to move away from a traditional "container" perspective of deprivation as defined by many existing composite deprivation measures so that it is considered as part of a dynamic process. The scale at which these traditional deprivation measures are created for do not provide sufficient specificity of local health disparities, and in an environment where PCTs must demonstrate best practice and value for money, social marketing campaigns derived from segmentation using deprivation indicators are unlikely to reach those with the greatest health needs. Geodemographics are applied at the smallest spatial unit available to UK researchers, the postcode unit (see table 6) which approximates 15 households, and because they are created by clustering together socially similar population sub-groups it is fair to expect them to provide a more useful differentiation of the socially similar characteristics.

As highlighted in the literature, health outcomes across all population groups and sub-groups in the UK have considerably improved over the last 100 years. But at the same time the nature of illness has changed. There has been a movement away from large epidemics of communicable diseases towards disease prevalence associated with behaviour choices and lifestyle (Hicks and Allen, 1999). Despite these improvements, there still remain different levels of disease incidence, morbidity and health outcomes across different sections of our society which is supported by a significant body of evidence (Marmot and Brunner, 2005). Missing from many of the indicators and measures developed is consideration for describing in detail, at the local level, materialist inequalities in health. By ascribing different neighbourhood Types with supplementary health data, information and knowledge of health outcomes can be acquired. As a consequence the research will attempt to identify variations across different social structures and the geographical scales at which they operate.

Despite the ever increasing number of qualitative and quantitative research projects, the contribution of social capital to understanding of health inequalities has not yet been absolutely defined and there still remains a lack of consensus. The literature identified a developing discussion surrounding the influence of neighbourhood differences in social capital and whether the lack of, or differences in, social capital across neighbourhoods has an influence on health outcomes (Subramanian et al., 2003; Lochner et al., 2003; Wakefield and Poland, 2005; Veenstra et al., 2005).

The contributions that the theoretical concept of social capital will make to the thesis arise from the notion that health-threatening behaviour is a response to material deprivation and stress (Jarvis, Wardle et al., 1999). This suggests an apparent dichotomous relationship between the material and psychosocial explanations of health inequalities upon which models of social determinants of health are built. Identification of likely health threatening behaviours for neighbourhoods will provide public health professionals with tools which can be used to build and harness social capital through social marketing and participation interventions through which communities are encouraged to change their behaviours by taking part in activities. Thus, if health threatening behaviours are measured at the level of the neighbourhood they may provide useful proxy measures of social capital. Existing measures of social capital are ill-defined due to the lack of consensus about what it actually means, but certain health-harming behaviours which are reinforced by social cohesion and social networks may provide new insights for identifying neighbourhoods and the presence of social capital. This is particularly relevant for the bonding social capital which relates to common identities and ties amongst similar people, where indicators of smoking behaviours may act as a useful proxy for predicating this form of social capital.

It is evident from the literature that there is a clear requirement to move towards developing measures that encompass community attributes based on social similarity. Aggregation of population sub-groups according to similarity will highlight gradients of health inequalities at the neighbourhood level and provide an enriched evidence base for public health interventions and campaigns.

The requirement for these new measures is further supported by the present lack of local neighbourhood measures of health needs and behaviours. Data protection issues restrict access and use of individual level records and mandatory requirements for relevant data reporting are not in place. One of the paradoxes of local health care needs assessment is data collected by general practitioners are not made available to Primary Care Trusts (PCTs) for secondary analysis. General Practices collect individual level data, but because they are private businesses they are not legally obliged to share data with the primary care organisations responsible for commissioning their services. The 1998 Data Protection Act and laws relating to medical confidentiality further ensure specific information is difficult to obtain.

4.3 Research aim and objectives

The strands of literature reviewed together with the existing research limitations identified in the problem definition have led to the formation of the following overarching aim for this thesis:

> **To develop and explore an alternative framework for measuring and disseminating health outcome indicators of inequality, to be used in social marketing and public health interventions.**

This thesis will present a more straightforward yet effective alternative to exploring the measurement of inequalities in health outcomes, by using geodemographic classifications to develop standardised neighbourhood risk indices. Exploration of these inequality gradients through social and geographical space will provide a new and detailed insight and provide greater understanding of local neighbourhoods for health practitioners and lay community members.

The development of an alternative framework for modelling inequalities according to the social similarity of neighbourhoods is justified for the following reasons. Alternative measures based on similarity move away from the traditional 'container' approach of measurement by administrative units and moves towards a modern social measurement basis for analysis; "social measurements translate observed characteristics of individuals, events, relationships, organizations, societies, etc into symbolic classifications that enable reasoning of a verbal or logical or mathematical nature" (Heise, 2001).

Such an approach should provide a useful and systematic method to the complexity of social differentiation.

Geodemographics provide measures that incorporate compositional and contextual influences on health need, which on completion of the thesis will realise:

a) An enriched picture of local neighbourhood health need beyond standard deprivation measures;
b) A framework for identifying neighbourhoods at risk;
c) A means of highlighting neighbourhood variation in likely health outcomes;
d) An alternative tool for understanding service equity (services reaching those with the greatest need);
e) Proxy indicators of social capital;
f) An intelligence led solution for social marketing campaigns;
g) A means of describing the health trajectories of small localities to explore how lifestyle may be represented using geodemographic indicators.

As yet, little use has been made of geodemographic indicators to differentiate population conditions of health outcomes, unhealthy lifestyles, behaviours that lead to poor health. Geodemographic classifications study population types and their behaviour as they vary by geographical area, typically clustering small areas by postcodes, combining variables from the census of population and other socio-economic data such as housing, financial and lifestyle information to classify neighbourhoods into different types. Residents are then assigned particular characteristics according to the neighbourhood Type within which they live.

The measurement of health needs and their associated outcomes forms a fundamental component of evidence-based policy, strategy and delivery of health care services at local scales, each of which aims to reduce overall inequalities. Exploring the use of a geodemographic based framework alongside national health datasets and surveys of population health, should facilitate the differentiation between local variations in health behaviours across neighbourhoods, enabling the development of expected measures/risk indices of health needs for local populations. This realises a requirement highlighted from the literature review, which should improve the specificity of social marketing campaigns in order to maximise efficiency.

Variability in health outcomes are often ascribed to two main causes. Compositional variables refer to the individual or the household e.g. age, gender or marital status. Contextual variables describe differences arising from a group, which cannot be reduced to an individual, for example absence of facilities in a neighbourhood; they can be associated to place effects of health inequalities. The framework and measures developed in the course of this thesis will explore the development of readily intelligible statistics that are comparable across neighbourhoods and to a national baseline that uses contextual neighbourhood variables to develop health need profiles for predicting some compositional variables.

To harness effectively the power of social capital through social marketing the new measures need to be accessible both to a range of public health professionals and lay communities. There is a need to ensure that community profile information for both communities and public health departments is accessible, informative, safe (anonymous) and easy to use (Rhind, 1992). Currently, limited information is available for communities to use and traditional indicators of inequalities and deprivation are firmly embedded in public health departments. A number of different approaches will be explored and evaluated. They will determine the most effective methods of communication, for imparting new tools and techniques in organisations, to improve the interfacing of research projects with routine work practices.

The theme of this research is to extend the health inequalities research and its associated data framework to explore variability in the spatial and social domain encompassing notions of social measurement and the identification of social facts related to health. This will be carried out by exploring the feasibility of a geodemographic framework for analysing health inequalities using the two tier method illustrated in Figure 11. The first tier will explore how a geodemographic framework can be extended to a variety of health datasets commonly used in health analysis. The second tier appraises this framework and explores appropriate methods of dissemination and sharing of best practice, through the evaluation of appropriate methods of communication to investigate the interface between research and knowledge transfer. This will enable the assessment of the effectiveness of turning research knowledge into professional and community knowledge.

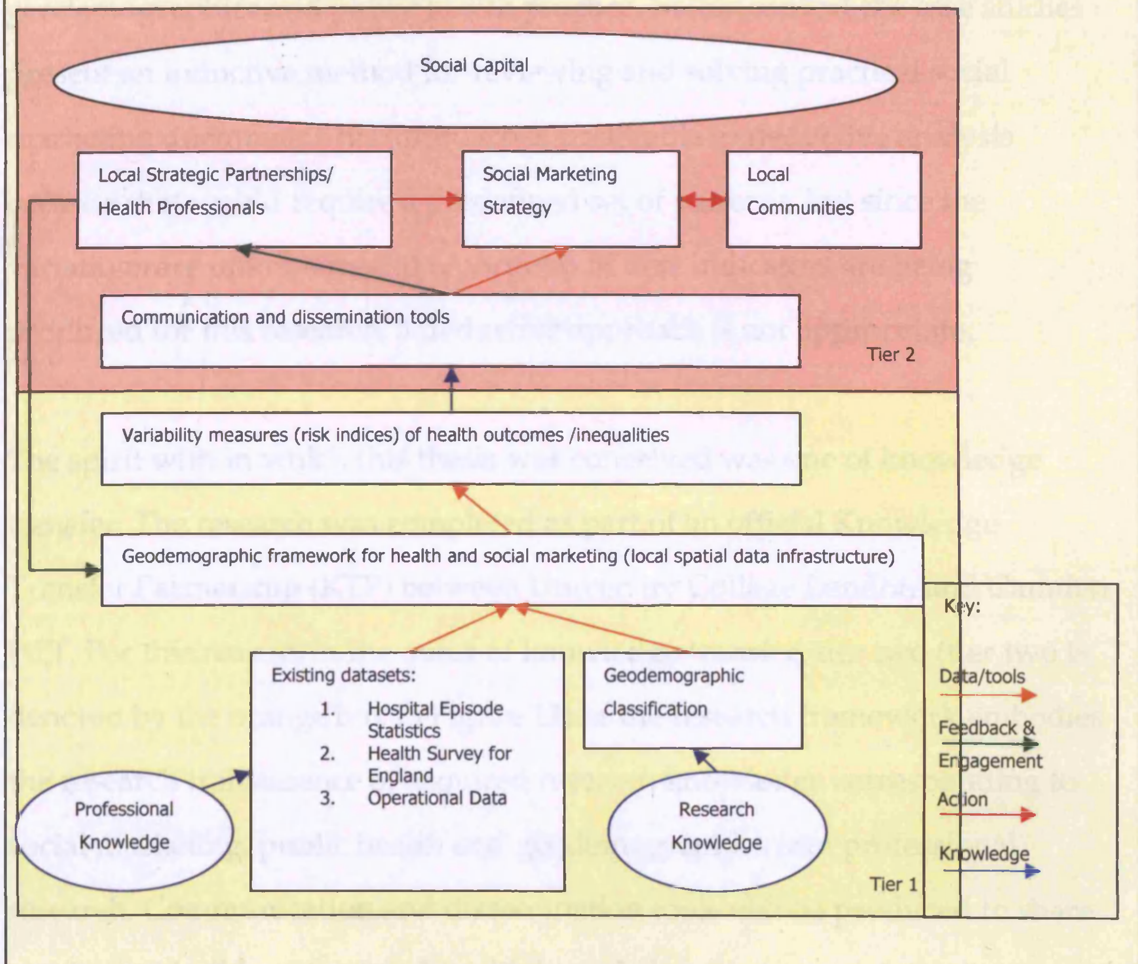


Figure 11: Research framework

In the first phase of the research (tier one is denoted by the yellow box in Figure 11) the organising framework for geodemographics in public health (Figure 13) will be adopted. Using an inductive approach, a geodemographic framework, through the development of a local spatial data infrastructure will be used to: (a) create new predictions of health risk and (b) explore and review these new health risk indicators across different neighbourhoods. On completion of Tier one of the research framework, new insightful health facts can be used to develop information about distributions of diseases of comfort among the local population, combining both professional health knowledge and research knowledge.

Tier one will be completed through the development of a series of case studies which slowly builds on the research knowledge in the field of geodemographics and public health practice. In this context the case studies present an inductive method for reviewing and solving practical social marketing dilemmas. This approach is preferable to deductive analysis because that would require a predefined set of patterns, but since the variations are unknowns and a portfolio of new indicators are being produced for this research, a deductive approach is not appropriate.

The spirit with in which this thesis was conceived was one of knowledge transfer. The research was completed as part of an official Knowledge Transfer Partnership (KTP) between University College London and Camden PCT. For this reason, in the quest of knowledge transfer, tier two (tier two is denoted by the orange box in Figure 11) of the research framework embodies the research transference of acquired research knowledge corresponding to social marketing, public health and geodemographics into professional research. Communication and dissemination tools will be produced to share results and build professional health knowledge.

In line with the research framework outline in Figure 11, a modular approach is adopted for the completion of this thesis, associated with the completion of each of the objectives set out below, as shown in Figure 12. Each module explores the potential of a practical GIS and geodemographic framework and a set of tools to explore the acquisition and diffusion of geodemographics for exploring health disparities and developing and harnessing social capital for public health use.

To fit with this modular, inductive approach three research objectives have been identified. On completion these objectives will contribute to the achievement of the underlying aim of this research. The three objectives are outlined in Figure 12 and directly line up with the structure presented in Figure 11. In Figure 12 the first two objectives fit with the first tier of the research framework, and the third objective fits with the second tier of the research framework. Each objective is discussed in more detail below.

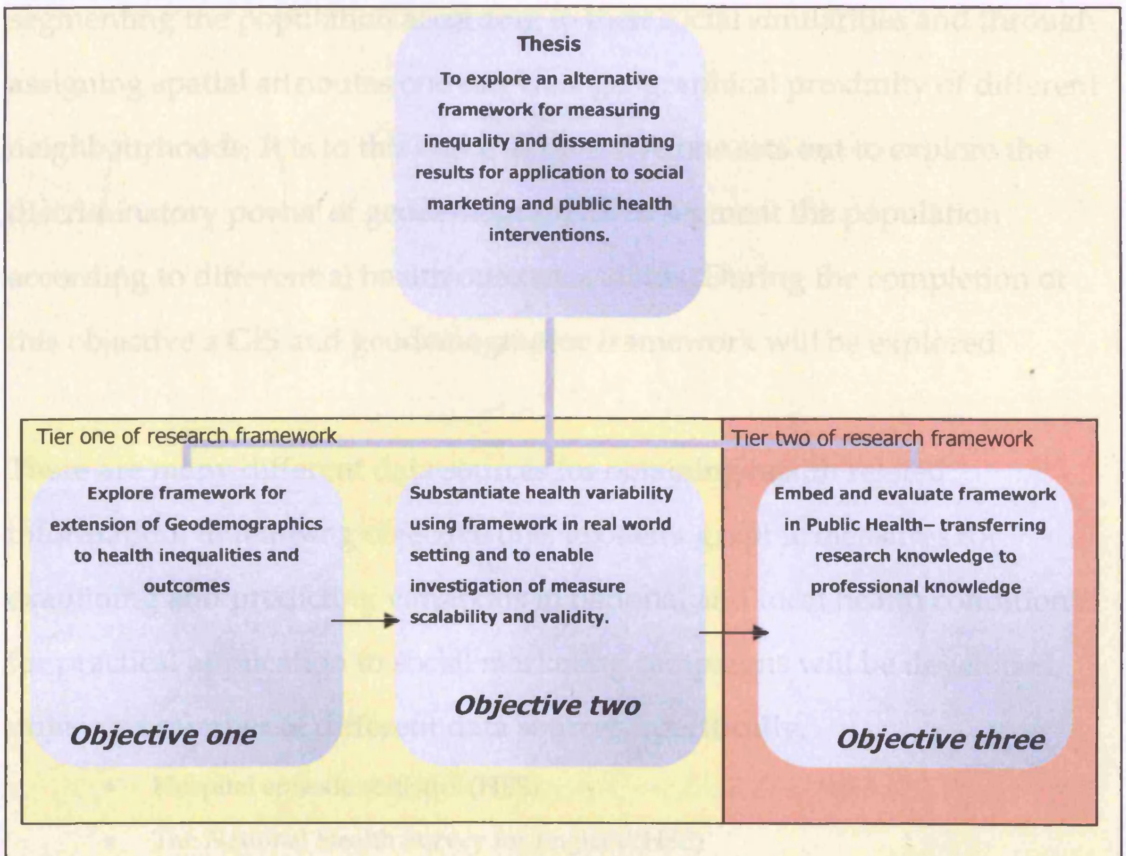


Figure 12: Outline of research aims and objectives

4.4 Objective one

Social marketing and health promoting interventions rely upon a comprehensive understanding of local populations to ensure their needs and requirements are fully recognised. As identified by Dedman and colleagues (2006, page 2), it is an intelligence led technique, requiring specific and robust data methods for identifying the most relevant population groups.

Locating population sub-groups and being aware of their motivations, cultural norms, social similarities and characteristics is a critical component of successful social marketing strategies (Kotler, 2002). Social marketing begins and ends with a focus on the individual within their social context (French and Blair-Stevens 2006). Geodemographics offers a mechanism for segmenting the population according to their social similarities and through assigning spatial attributes one can view geographical proximity of different neighbourhoods. It is to this end that objective one sets out to explore the discriminatory power of geodemographics to segment the population according to differential health outcomes/status. During the completion of this objective a GIS and geodemographic framework will be explored.

There are many different data sources for obtaining health related information. In realising objective one, geodemographic measures for examining and predicting variations in national and local health conditions for practical application to social marketing campaigns will be developed, utilising a number of different data sources, specifically:

- Hospital episode statistics (HES)
- The National Health Survey for England(HSE)
- Operational health data, such as general practice(GP) patient lists.

These datasets represent some of the data available to health researchers in England. A detailed description of each is documented in the next chapter. The hospital episode statistics (HES) are derived from the national administrative database that records information about all people who are admitted to hospital (not outpatients or accident and emergency). By comparison the health survey for England (HSE) is an annual survey which uses qualitative surveying techniques of England's population providing a health of the nation report. Other operational data available to researchers, with appropriate access, are GP patient lists. These datasets are discussed in greater detail in section 5.3. By combining these datasets with the geodemographic classification it will be possible to create new, enriched local descriptors of health.

4.5 Objective two

The term 'neighbourhood' in the context of this research refers to the immediate residential environment of individuals and their shared geographic and social space, measured using the postcode unit (see section 5.3.1). They are frequently characterised by a number of different factors, such as physical characteristics, social and economic resources, population interaction. Geodemographics defined by the neighbourhood (often based upon UK unit postcodes) will be used to substantiate the methodology presented, to explore the variability of the measures developed in objective one and evaluate their scalability and validity in relation to establishing whether they are fit for purpose. In the context of real world problems faced by public health professionals, the measures will be evaluated to determine if the neighbourhood health indices are effective and specific enough to solve the complex challenges faced.

4.6 Objective three

The final objective of the research attempts to evaluate the way in which the indicators can be embedded into existing health promotion and prevention initiatives. The framework and measures developed during the course of completing objectives one and two should be disseminated and best practice shared with public health practitioners. Objective three will explore and evaluate a number of alternative mechanisms and tools for communicating and distributing appropriate information to different audiences. The chapter first adventures into a brief discussion on the wider issues related to technology diffusion in the NHS before developing a framework for implementing geodemographics in the PCT, using a web-based intranet tool.

4.7 Chapter summary

Since the re-emergence of geodemographic classifications in the public sector, very few studies in the health domain have taken advantage of the new local postcode classifications of neighbourhoods. The benefits of utilising such classifications for social marketing have yet to be realised.

There are four general needs for improving the methods of identifying health inequalities across the social spectrum. The needs of which are: (a) a need for an intelligence led method for developing neighbourhood social marketing campaigns; (b) a need for a more sophisticated approach for identifying health risk that goes beyond indices of deprivation and cruder health indicators; (c) a need to supplement standard geodemographic classifications with specific health data to leverage the richness of information that profiling provides; (d) the need to provide health professionals with a vehicle for which to adopt new advances in health needs prediction. These needs identify some of the gaps in the literature which this thesis sets out to address.

These four needs will be realised through the adoption of a multi tier approach (as described in Figure 11) because it is well suited to knowledge transfer. The modular case study approach means that the short term benefits befitting results dissemination results in evidence and knowledge can be transferred quickly. The forthcoming chapter, the final one in this section, takes the research objectives identified in this chapter and sets out the research method that will be used in section II of the thesis to complete both sections of the research framework.

Chapter 5

Research Methodology

5 Research methodology

It is common practice for academic research projects to occur in isolation and exploitation of research findings and best practices in local government sectors are beset by many obstacles. Consequently, within local government the adoption of new and innovative techniques and tools may often be slow. An objective of the research is to provide a foundation for public health practitioners and community members to explore GIS and geodemographic analysis for informing social marketing interventions. This will share the knowledge and best practice that was built up. This chapter describes the research method used to complete the three main objectives described in Chapter 4. This chapter considers the type of data available, how they can be integrated to form a spatial infrastructure and notes the key processes and techniques used to develop the forthcoming measures. The chapter closes by introducing the reader to the main study area of Camden, London, UK. Camden was chosen as the research laboratory because the Public Health department of Camden Primary Care Trust were co-sponsoring this research.

The research methodology comprises a number of interrelated sections. First, key definitions and concepts are described. These sections are followed by a description of the resources required, and the methods of data collection that will be used. Finally some of the high level processes and techniques are documented. In the next section case studies are conducted in a public health setting in order to demonstrate the real world application of the method. Details of the most applicable methods to be used with each dataset are documented in an extended discussion of the case studies (Chapter 8). Issues relating to the methodology are discussed in the data dissemination chapter of this thesis (Chapter 9), which also discusses attempts to mainstream the method and disseminate techniques into public health practice. Some reactions to the method by public health departments will also be explored in Chapter (9) about the results dissemination.

5.1 Understanding the potential of geodemographics

With the emergence of new technologies and improving data quality, it is now possible to understand population differences at a number of different scales; local, regional, national and even international. These developments have greatly advanced the discriminatory power of geodemographic classifications. During the late 1970s classifications were only available for Census wards but since the availability of Census data for output areas (OA) and enumeration districts (ED) taxonomies of populations have been created for smaller scale local neighbourhoods and even households.

Geodemographic analysis in conjunction with spatial data can assist the strategic delivery of public health priorities. Their application is not affected by issues associated with data protection and confidentiality. They have the potential to provide an alternative evidence base for decision makers to understand local differences within the realms of social, economic and health disparities. Geodemographic tools and techniques have been used extensively within the commercial sector for many decades. In 1995, Leventhal (page 223) illustrated three main applications of these classifications; survey design, retail planning and direct marketing. Since then Harris et al (2005, Chapter 5) adapted an existing model of geodemographic applications for commercial applications from Curry (1993 page 200). Using the research by Harris et al (2005) and earlier work of Leventhal (1995) it was possible to develop an organising framework for the conceptualisation of the geodemographic framework developed throughout this thesis and for building the case studies.

These models have led to the development of the model of health applications of geodemographics shown in Figure 13, based on the local public health priorities identified in 2005 (the time this research began) for the London borough of Camden, the case study area for this thesis (see section 5.6). The diagram clearly identifies a range of applications for geodemographics in community and health activity profiling; strategic resource provision, service equity and needs assessment, and comparison based decision making. Each of the applications identified can be used to inform social marketing campaigns or more traditional community health interventions. Examples can be seen in the case studies conducted during this thesis.

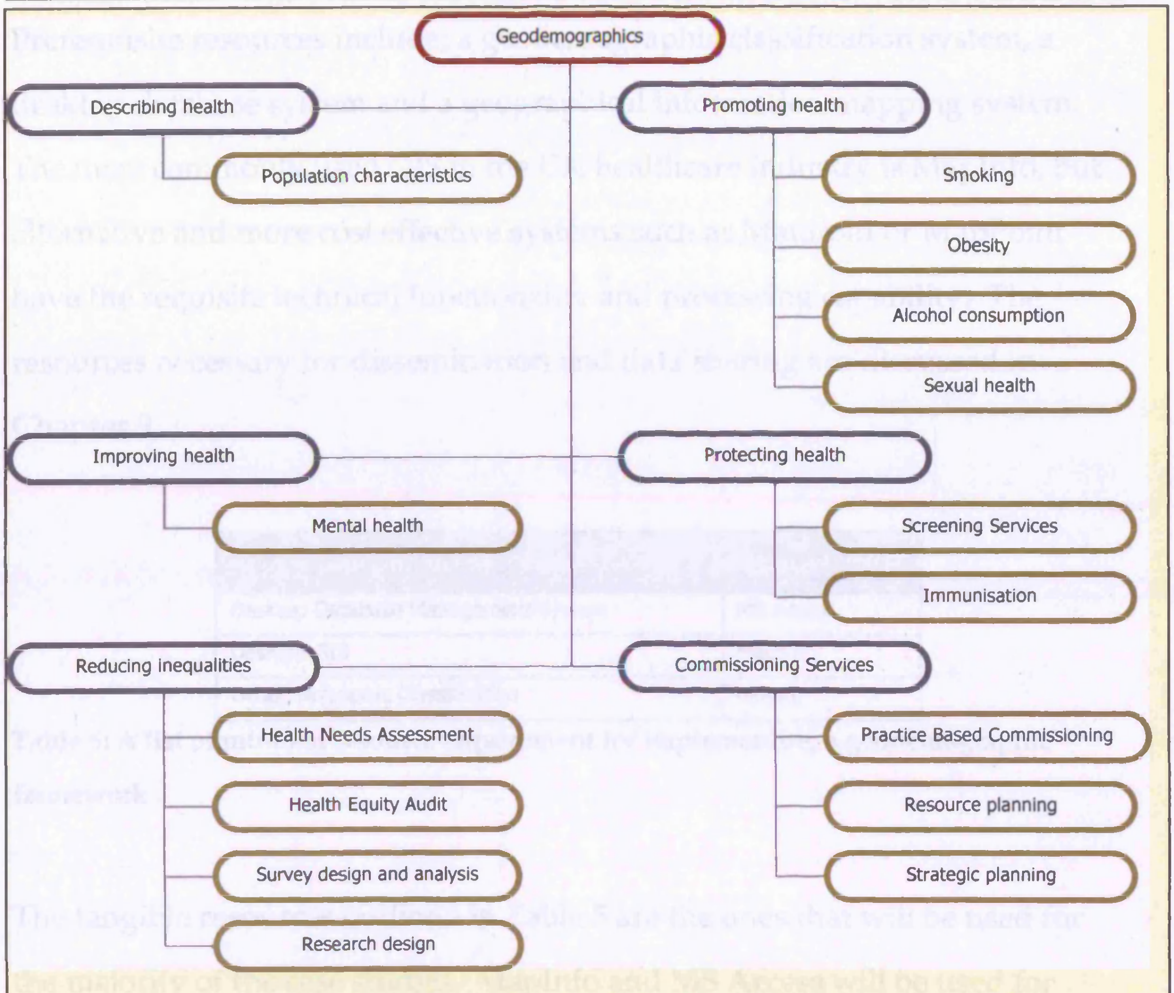


Figure 13: Public health applications of geodemographics

The diagram above provides a comprehensive summary of the potential public health fields where a geodemographic framework would provide an enriched evidence base for decision making. In essence it provides the organising framework for the case studies of the thesis.

5.2 *Resources required*

In order to implement a geodemographic framework, a number of essential resources are required. This section outlines the resources necessary for implementing a low-tech, low budget implementation that makes best use of everyday desktop technologies and software, chosen to fit in with the need to disseminate results in a resource constrained environment (see Chapter 9).

Prerequisite resources include; a geodemographic classification system, a desktop database system and a geographical information mapping system. The most commonly used GIS in the UK healthcare industry is MapInfo, but alternative and more cost effective systems such as Manifold or MapPoint have the requisite technical functionality and processing capability. The resources necessary for dissemination and data sharing are discussed in Chapter 9.

Resources	Example
Desktop Database Management System	MS Access
Desktop GIS	MapInfo
Geodemographic Classification	Mosaic

Table 5: A list of minimal resource requirement for implementing a geodemographic framework

The tangible resources outlined in Table 5 are the ones that will be used for the majority of the case studies. MapInfo and MS Access will be used for geographical analysis because software licenses had been previously bought by Camden PCT, the PCT which co-sponsored this research along with the EPSRC. The geodemographic classification Mosaic is used here because of existing academic links with the system designers.

5.3 Data acquisition

Data acquisition is an important part of any research method or knowledge base. The quality of the processing, analysis and results is ultimately determined by the validity of the input data. There are many types of data that can be used within a geodemographic framework. Those that will be used to complete the case studies have been outlined in the section below.

5.3.1 Geographical boundaries

5.3.1.1 United Kingdom administrative boundaries

Official statistics produced by local and regional government are commonly aggregated to the following geographies: output area (OA), super output area (SOA), ward or local authority (see Figure 14 and Table 6). The 2001 Census OAs were designed with statistical analysis in mind, representing a paradigm shift away from traditional census zoning based on political or organisation considerations (Martin, 2002). They were built automatically using consistent and systematic criteria building upon the Ordnance Survey Address Point dataset, and postcode units polygons created by Thiessen polygons (at that time of designing the 2001 Census, there was no boundary data available for the postcode unit point dataset which is why efforts were concentrated to design some).

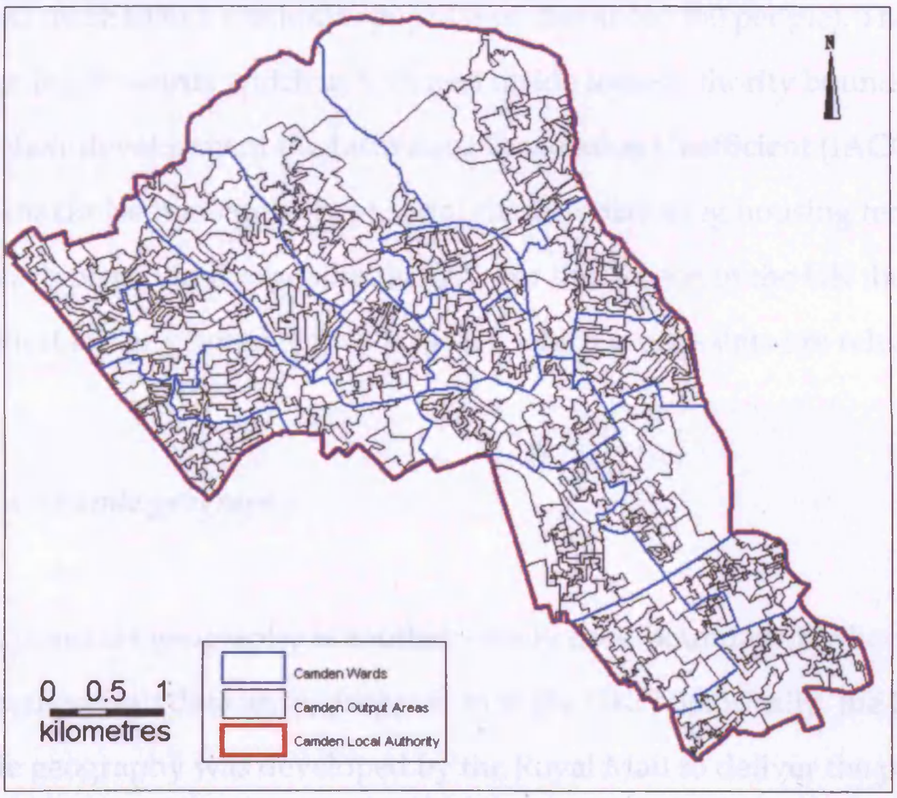


Figure 14: Common UK administrative units used in local government

Geographic unit	Number in UK - September 2004	Population
Output Area (OA)	175,434 in England	Minimum OA size is 40 resident households and 100 resident persons but the recommended size was rather larger at 125 households (ONS, 2006).
Super Output Area (Middle Layer)	6780 in England	Min 5000 Mean 7200
Ward	7932 in England	Varies
Local Authority	434 in UK	Varies, e.g. in Inner London ~ 200,000

Table 6: Administrative geography of the UK

The algorithm used to create the OAs for the 2001 Census applied rules that constrained postcode polygons to statutory boundaries and topographic features, splitting postcode polygons where they crossed ward boundaries and ensuring a nested geographical hierarchy was created. The zoning procedure was further constrained by the need to ensure all output areas contained more than a minimum population threshold (50 people). The 2001 OAs nest inside wards which in turn nest inside local authority boundaries. During their development the Intra Area Correlation Coefficient (IACC) was used to maximise homogeneity of social characteristics (eg housing tenure) within each postcode comprising the OA. For this reason in the UK they are the smallest socially homogenous areas for which census data are released.

5.3.1.2 *Postcode geography*

The UK postcode geography is another widely used boundary file because it is the smallest unit data are aggregated to in the UK. Historically, the UK postcode geography was developed by the Royal Mail to deliver the post. It is a summary of an address in a form which can be read by the computer and enables the automatic sorting of post (Raper et al., 1992). Almost everybody knows their postcode, which is why the Office of National Statistics uses it as their main geographic reference for data collection. This reference may then be linked to the common administrative boundaries using aggregation techniques (Office of National Statistics, 2005). The postcode is now commonplace in geographical referencing in the UK and is used as an identifier in many different population registers.

Use of this method of georeferencing is subject to a number of caveats. The postcode geography invariably do not always nest within wards, because some had to be split across boundaries, and thus users must decide how to allocate attributes associated to postcodes between areas. Originally unit postcodes were designed as part of the postman's walk and do not necessarily correspond appropriately to social or physical environmental boundaries. They are limited in the number of addresses they can contain. In 1992 only 3% of postcodes had 50 delivery points or more (Raper et al., 1992, page 38). Also whilst recoding existing postcodes is generally not practiced it cannot always be avoided (Raper et al., 1992, page 39). The structure of the UK postcode geography is outlined in Table 7. In the context of this research a unit postcode is taken to approximate to a neighbourhood boundary, the functional form of social space.

Example	Geographic unit	Number in UK - September 2004	Number of Households
N	Postcode Area	124	
N1	Postcode District	2934	
N1 3	Postcode Sector	9903	
N1 3TR	Postcode Unit	1.7 million	~12 to 16

Table 7: Postcode geography of UK

5.3.2 Health data

There are a number of different datasets available for health inequalities research. The most commonly used datasets are official records of births and deaths, Census data and hospital episodes statistics (HES): see Table 8. With respect to operational health data (births, deaths and hospital episodes etc), the level of aggregation available depends upon the level of user access restrictions. For example hospital episode statistics at the scale of individual patients associated with postcodes are available for secondary analysis by NHS staff and those with honorary contracts, but are aggregated to super output areas for uses external to the NHS.

The HES data is a rich information source. It is a national database of statistics for the care provided by hospitals in England. It contains records of admissions for every patient episode in England. There are limitations surrounding its use. Originally the dataset was set up to manage the financial costs of admissions, rather than for disease monitoring and prevention. Perhaps as a consequence, not all personal data are collected accurately. Also the dataset only includes conditions that are serious enough to require admission to hospital and therefore represents use rather than actual health need or demand.

Disease registers of diabetes, heart disease, and obesity are held internally within general practices but are only shared with PCTs following aggregation to practice level for performance monitoring purposes. The only available resource at the scale of the individual is the patient register, which records basic information on individuals (eg age, sex, place of birth and postcode) and is managed by PCTs and maintained by general practices. Difficulties arise for Primary Care Trusts, which are mandated to provide and commission appropriate services (DoH, 2004) to meet the health needs of populations because they have little detailed patient information relating to health needs and inequalities. This results in a need to develop proxy indicators using survey data and geodemographic data (Weber 2004).

Name	Source	Analysis Unit available to this project	Key Features	Description
Camden Registered Patients	Camden PCT	Postcode Unit	Age, Sex, Place of Birth, NHS number.	List of patients who are registered with Camden general practices.
Screening Data	Camden PCT	Postcode Unit	Uptake of Services.	List of all patients who have been called for NHS screening programmes.
Hospital Episode Statistics	Camden PCT	Postcode Unit	Diagnosis, Age, Sex, Ethnicity, Postcode.	Database containing information about each hospital episode for every patient admitted to hospital in England.
Births	Camden PCT/ ONS	Postcode	Sex, age of Mother, postcode.	Register of births in the UK.
Deaths	Camden PCT/ ONS	Postcode	Age, sex, postcode, cause of death.	Register of deaths in the UK.
GP location	Camden PCT/ ONS	Address Point of Postcode	Location of practices.	Point file of general practice locations, number of doctors and their genders.
Census Data	ONS	Output Area	Self defined health status.	10 year survey of all people and households in the country.
IMD 2004	Office of the Deputy Prime Minister	Super Output Area	Composite measure of deprivation.	Seven domains relating to: income, employment, health and disability, education, skills and training, barriers to housing and services, living environment and crime.

Table 8: Health datasets available to this research

5.3.3 Survey and lifestyle datasets

There is a considerable lack of detailed information about lifestyle characteristics available to health inequality researchers. Thus researchers have turned to alternative data sources to produce estimates, such as those shown in Table 9. The Health Survey for England (HSE) is the most extensively used public sector survey. Annually 19,000 people are surveyed across England to obtain information about the health of the nation. The focus of the survey changes each year, but there are core demographic and lifestyle elements consistently asked each year. The survey is carried out jointly by the National Centre of Social Research, London and the Department of Epidemiology and Public Health at University College London. The sample size becomes too small if analysis is focused upon small neighbourhoods therefore its general release is available and statistically valid for large geographical areas, such as London boroughs. The survey asks responders questions about obesity, their drinking patterns and smoking habits; variables with a close association to diseases of comfort. The survey is used in the chapter 6 to explore neighbourhood patterns of survey respondents and further details will be provided there.

Lifestyle data are also available alongside many commercial geodemographic classifications. The lifestyle information in these systems is derived from a number of sources, including the Target Group Index (TGI lifestyle survey). This survey provides information that describes groups of people and their consumption of consumer goods and services based on 4000 brands, 500 product categories and 250 lifestyle statements. This survey is not directly used in this research, because it is already commonly used by marketers and is incorporated into commercial geodemographic variable datasets. This survey was excluded because of the heavy focus on consumption behaviours resulting in questions often commissioned by companies that are brand specific.

Name	Source	Analysis Unit	Key Features	Description
Mosaic - Geodemographic classification	Experian	Postcode	Global index contains 1000 lifestyle variables	Small areas measure of socio-economic and lifestyle. They are created by clustering over 400 different variables according to social similarity.
TGI - Lifestyle Survey	Richard Webber/Experian	Postcode	Diet and lifestyle variables	Target Group Index (TGI) consumer survey that annually samples 19,000 people within the UK population to produce consumption profiles.
Health Survey England	NATCEN	Publicly available for PCT boundaries only	BMI, Drinking, Smoking	Annual Health Survey for England, samples the health characteristics of 19,000 people. Standard questions remain the same each year – but key focus changes annually

Table 9: Survey and lifestyle datasets for measuring and monitoring health conditions

There are many commercially developed geodemographic typologies. The one used in this study is called Mosaic, developed by Experian. Mosaic has been designed to identify different groupings of the population based on attribute similarities. It provides a postcode level classification of different types of people living in different neighbourhoods, based on similarities in income, education, and household type, as well as attitudes and lifestyles preferences enabling study of population types and their behaviour.

The commercial geodemographic classification used in this study was the 2004 version of Mosaic (Experian, Nottingham, UK). It was used because of existing academic links. This particular geodemographic classification clusters together over 400 different variables, according to social similarity and proximity, using the processes described in chapter 3.3.2. Over 400 different variables are used to create the Mosaic geodemographic typology:

sources include the 2001 census and other available data sources including the edited Electoral Roll, Experian Lifestyle Survey information, and Consumer Credit Activity, alongside the Post Office Address File, Shareholder Register, House Price and Council Tax information. The result is a classification of the 1.7 million UK residential postcodes into 61 different Types. The different neighbourhood Types are then arranged into a hierarchy of 11 Groups. A breakdown of the geodemographic population for the study area is included in section 5.6. To ensure the reader does not become confused any reference specifically to the Mosaic geodemographic Types or Groups the words will be capitalised, while general discussion of population groups and sub-groups will not be.

5.4 Data integration and management, a local spatial data infrastructure

The previous section identified the myriad of disparate datasets that will be used to create the case studies and to implement the research methodology. Careful data integration and manipulation is required to successfully utilise them. In order to create an appropriate evidence base for public health, it is critical to carry out the integration of datasets in a careful and logical manner.

Data integration and management is an important component of any geodemographic analysis. Figure 15 outlines the conceptual data model that will be implemented. There are a variety of datasets that must be integrated and the method for achieving this will depend upon their spatial resolution. Lifestyle data and the geodemographic classification are available for postcode units. For integration with other datasets either the grid reference of the postcode (as defined by Gridlink or CodePoint) will be used or the neighbourhood Type/Group associated with the postcode will be the identifier. Health survey data, hospital episode statistics, and screening data will be ascribed their neighbourhood characteristics by appending the postcode unit of the records to the geodemographic classification (to ensure anonymity in the data) and joined accordingly.

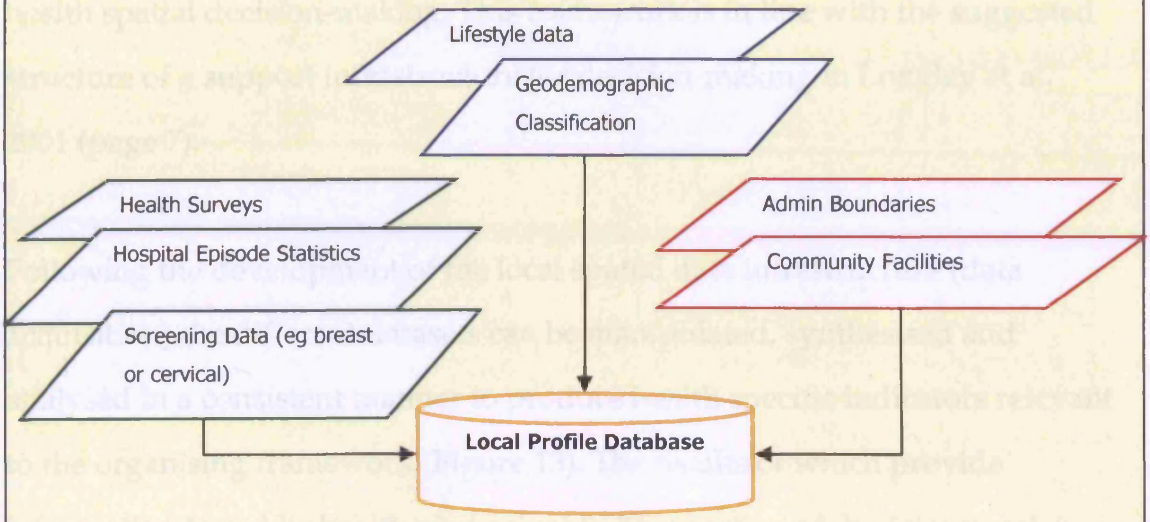


Figure 15: Conceptual data model

5.5 Processes and techniques

A number of processes will be used for the development of geodemographic health. The more detailed ones will be handled specifically in the context of each case study, but the generic processes crossing all the cases will be briefly document in this section. The key processes and development framework are outlined in Figure 16, which shows the progression from raw input data through data processing to the final output and results dissemination; whereby the results are communicated and provided for everyday use within the public health department of Camden PCT.

The framework outlined in Figure 16 is comprised of two distinct parts. The first part corresponds to processes associated with data acquisition, processing and analysis. The component processes of this part are responsible for transforming individual facts and data into information which in the second part on the model can be consolidated into evidence and knowledge. Here the conceptual data model (Figure 15) is turned into a local spatial data infrastructure for Camden, which provides the vehicle for informing public health spatial decision-making. This framework is in line with the suggested structure of a support infrastructure for decision-making in Longley et al. 2001 (page 7).

Following the development of the local spatial data infrastructure (data acquisition) the different datasets can be manipulated, synthesised and analysed in a consistent manner to produce health specific indicators relevant to the organising framework (Figure 13). The results of which provide information to public health professionals. This section of the framework is addressed in Chapter 6 and 7 of the thesis.

Thus far the skills knowledge has been primarily in the hands of the researcher (see Figure 11). The second part of the processing and development framework (Figure 16) corresponds to data sharing and dissemination of results. This section of the framework is responsible for transferring the research knowledge into the professional domain. It is here that the information created is turned into public health knowledge. Chapter 9 considers this section of the model and sets out to address the development of geodemographic knowledge transfer.

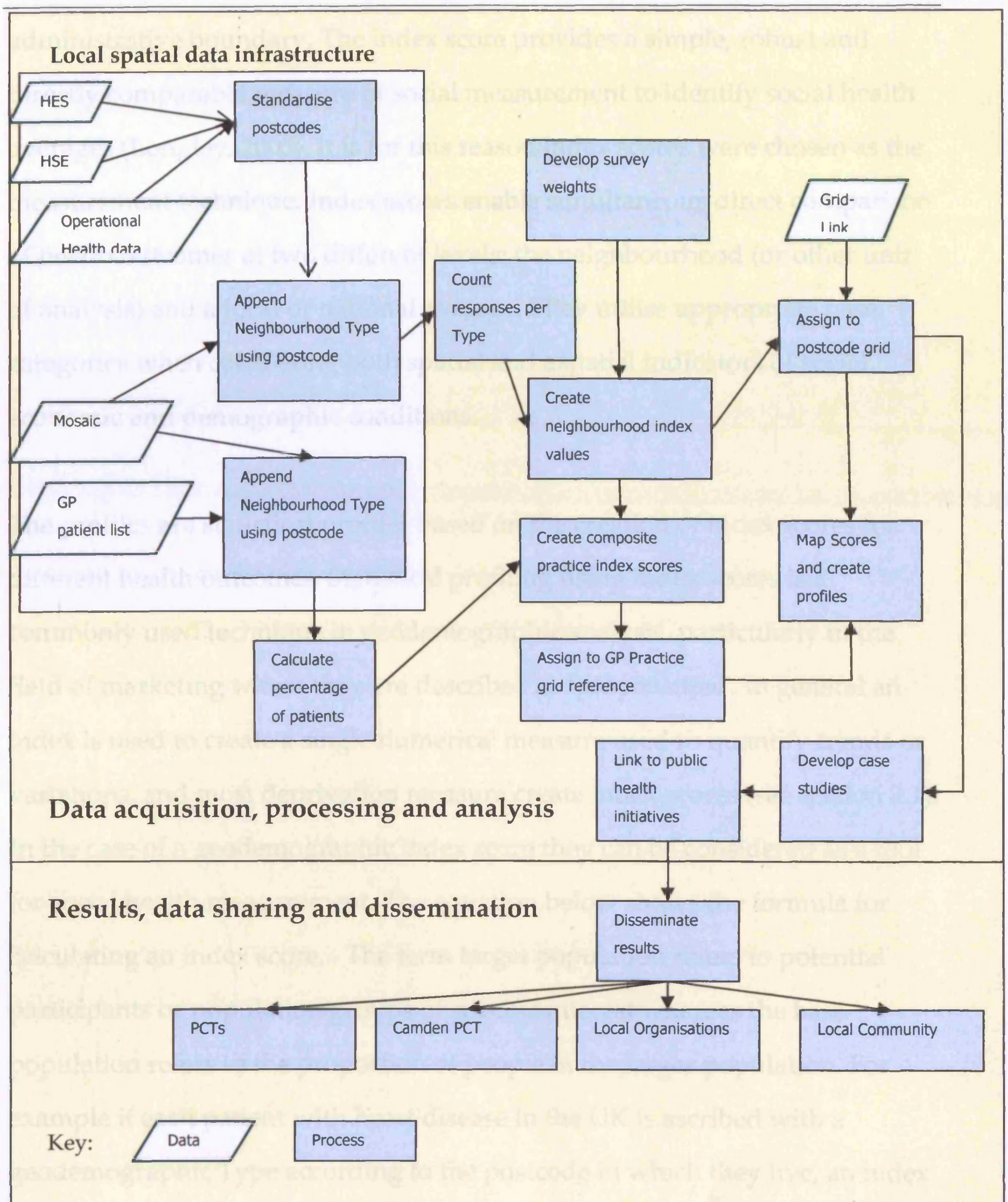


Figure 16: Process diagram and geodemographic development framework of the research design

Once the local spatial data infrastructure has been developed, health outcome indicators can be created to profile population sub-groups. In the context of this research, profiles are created for different types of analysis units representing different social and/or spatial scales; the neighbourhood (postcode unit), general practice (GP) and Primary Care Trust (PCT) administrative boundary. The index score provides a simple, robust and directly comparable measure of social measurement to identify social health averages (Longley, 2005). It is for this reason index scores were chosen as the measurement technique. Index scores enable simultaneous direct comparison of health outcomes at two different levels; the neighbourhood (or other unit of analysis) and a local or national average. They utilise appropriate base categories when calculating both spatial and aspatial indicators of social, economic and demographic conditions.

The profiles are statistical profiles based on the creation of index scores for different health outcomes. Statistical profiling using index scores is a commonly used technique in geodemographic analysis, particularly in the field of marketing where they are described as 'propensities'. In general an index is used to create a single numerical measure used to quantify trends or variations, and most deprivation measure create index scores (see section 3.1). In the case of a geodemographic index score they can be considered as a tool for social health measurement. The equation below shows the formula for calculating an index score. The term target population refers to potential participants or population groups of specific interest whereas the base population refers to the proportion of people in the larger population. For example if each patient with heart disease in the UK is ascribed with a geodemographic Type according to the postcode in which they live, an index score can be developed to predict the likelihood for all people living in particular neighbourhood Type in the UK compared to average propensity for the UK. The proportion of target population would be created by calculating the total count of people in the neighbourhood Type with heart disease versus the total number of UK residents living in that particular neighbourhood Type. The proportion of the base population would calculate the proportion of the total population with heart disease in the UK resident population.

$$\text{Index score} = \frac{\text{Proportion of target population}}{\text{Proportion of base population}} \times 100 \qquad \text{Equation 1}$$

The index score provides a simple, robust and directly comparable measure of social measurement to identify social averages. They can be created for health data to predict health outcomes by ascribing the geodemographic characteristics of a neighbourhood to the patient or survey respondent through the use of the postcode unit. This is a simple six step process, outlined in Figure 17. For each respondent/patient there are data records containing information about their address with the full unit postcode (step 1). The responders/ patients are the target population. The postcode of each respondent/patient is extracted and used as a unique identifier (step2). Using this postcode, the geodemographic classification is assigned to each postcode in the target population (step 3). The number of responders/patients for each geodemographic Group/Type is totalled (step 4) and the index value is then created (step 5), using the equation above. The final step involves the analysis of index values (step 6), which will be considered in the next chapter.

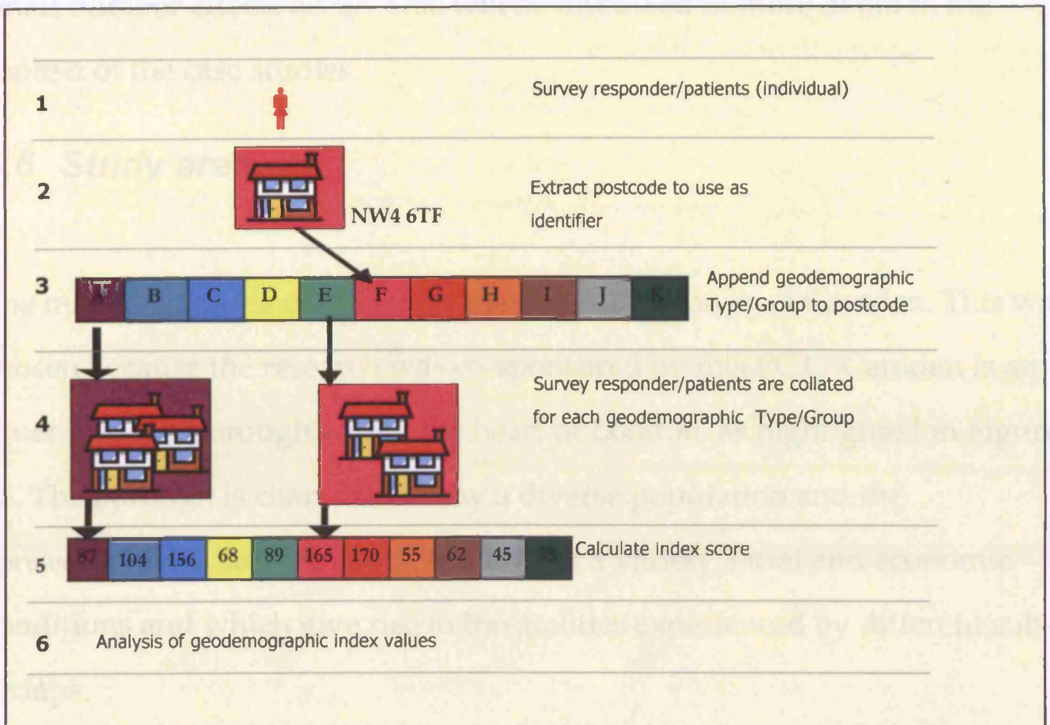


Figure 17: Method for calculating a neighbourhood (geodemographic) index, using operational health or survey data

In essence an index score is used to compare, for example, the number of diabetics in an observed population sub-group (target population) as compared to the total number of diabetics in a population base group (base group). Essentially observational frequencies of behaviour, presence of a disease or a survey response are compared to the total frequencies in a base population or survey sample. They are useful measures for identifying differences between sub-groups, but are not always useful for identifying observational differences actually within Groups/Types. There are a number of precautionary notes to be aware of when using index scores, some of which are documented by Harris et al (2005, page 121). The scores are subject to scaling effects; using a value of 100 to act as the average value means scores below 100 are constrained to 0, but scores above 100 can theoretically stretch to infinity. This creates particular issues on datasets that have very small sample sizes, because the resultant scores would become less reliable as small-number effects occur. This will be discussed in more detail in the context of the case studies.

5.6 Study area

The main study area of interest is the London Borough of Camden. This was chosen because the research was co-sponsored by this PCT. Camden is an Inner London Borough within the heart of London, as highlighted in Figure 18. The borough is characterised by a diverse population and the consequential polarities which result from a variety social and economic conditions and which give rise to inequalities experienced by different sub-groups.

Figure 18: Location of study area, London Borough of Camden

The 2001 Census showed that 198,020 people were resident in the Borough, although this has since been estimated to have risen to over 200,000. Of this population, 43% of residents were under the age of 30, a typical pattern for metropolitan areas which invariably have a much younger population pyramid. Almost half of the residents in full-time employment were educated to degree level, and what is more Camden has the largest proportion of full-time students living in it compared to the rest of London (11%). It is also home to University College London, located in the south of the Borough.

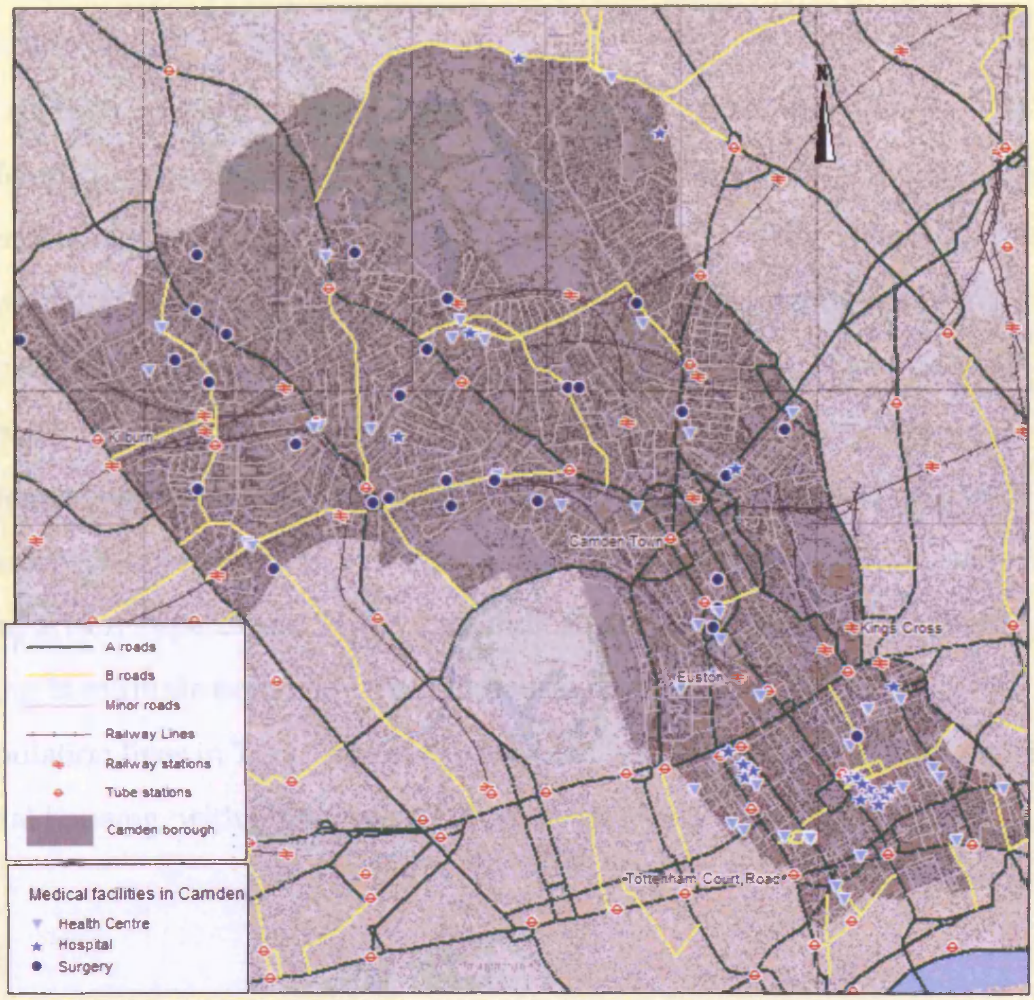


Figure 19: Map of Camden's medical facilities and local infrastructure

Figure 19, highlights the medical facilities and local infrastructure of the borough. It is home to 14 train stations and 20 tube stations and most recently became home to the new international Eurostar terminal at King's Cross. Around the main train stations of King's Cross and Euston are some very marginalised communities. The borough is resident to a multicultural society with 27% of the population from black and ethnic minority communities; with the Bangladeshi being the largest of these cultural communities.

5.6.1 Neighbourhood composition of Camden

On a national scale the postcodes are classified (using Mosaic) into 61 different Types and these are arranged into 11 different Groups. Within Camden, of the 11 national Groups 7 can be found and of the 61 national Types, 21 are represented in Camden. The four most frequently occurring Types are: Group A - Type 01, Group E - Type 28 and Type 29 and Group F - Type 36. Each of the characteristics represented by these postcode types differ considerably. People living in Type 01 postcodes are most likely to be career professionals living in sought after locations, with high incomes. By comparison Type 28 and Type 29 are indicative of areas with transient singles living in multiple occupied large old houses. Finally, 25% of Camden's population lives in Type 36 postcodes, which is representative of high density social housing, with high levels of diversity and very low incomes.

Mosaic UK Type	Population of Camden in Each Type	Percentage Camden Population	Index for Camden versus UK
A 01	51032	22.49	4057
A 02	10603	4.67	511
A 03	667	0.29	23
B 08	1780	0.78	50
D 26	259	0.11	10
D 27	1533	0.68	42
E 28	57716	25.44	2323
E 29	28206	12.43	1348
E 30	4768	2.10	178
E 33	3955	1.74	228
E 34	3549	1.56	317
F 35	1289	0.57	127
F 36	56865	25.06	1577
F 38	1585	0.70	235
F 39	1391	0.61	68
F 40	99	0.04	10
I 48	478	0.21	44
I 49	364	0.16	12
I 50	185	0.08	9
J 51	82	0.04	8
J 52	502	0.22	22

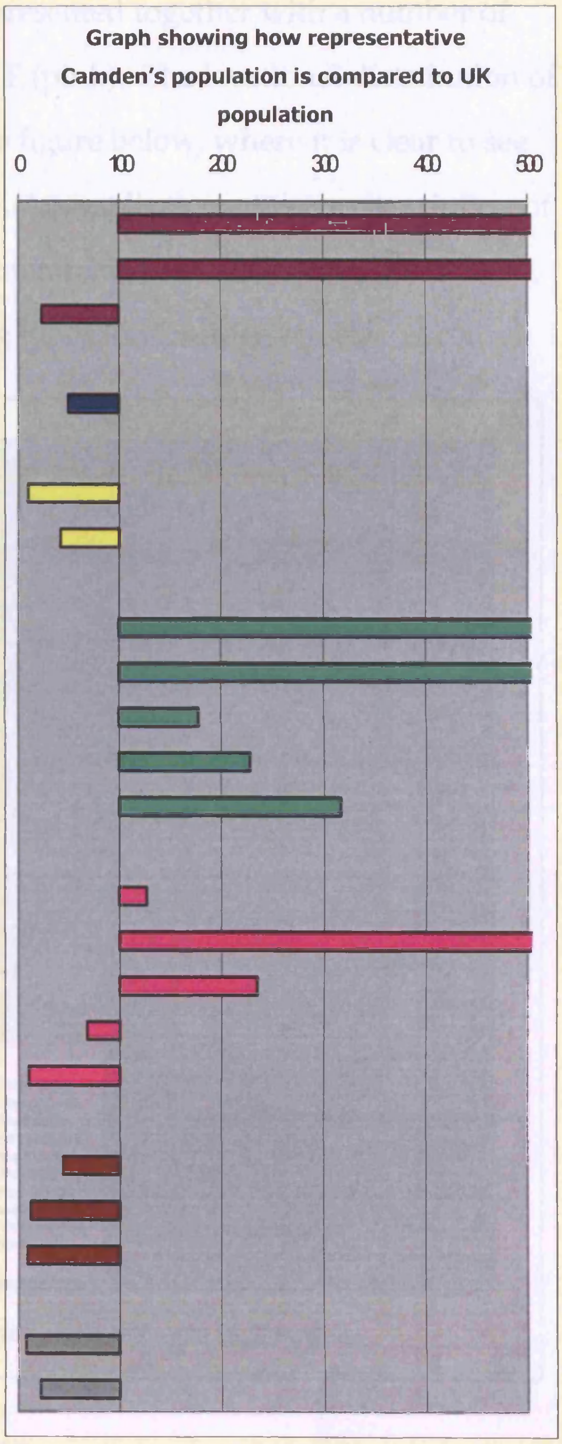


Figure 20: Neighbourhood Types represented in Camden

The graph in Figure 20 visually represents the population of Camden compared to the distribution of population Types in the UK. From this it is possible to see which neighbourhood Types are over or under-represented in the population. From this graph it is possible to see all of the Types in Group

E (colour coded – green) are over represented together with a number of Types in Group A (dark purple) and F (pink). The locational distribution of neighbourhoods is represented in the figure below, where it is clear to see Camden's great population diversity. Appendix A contains a description of the neighbourhood Types and their hierarchical Groups with photos representing their associated housing types for Camden streets.

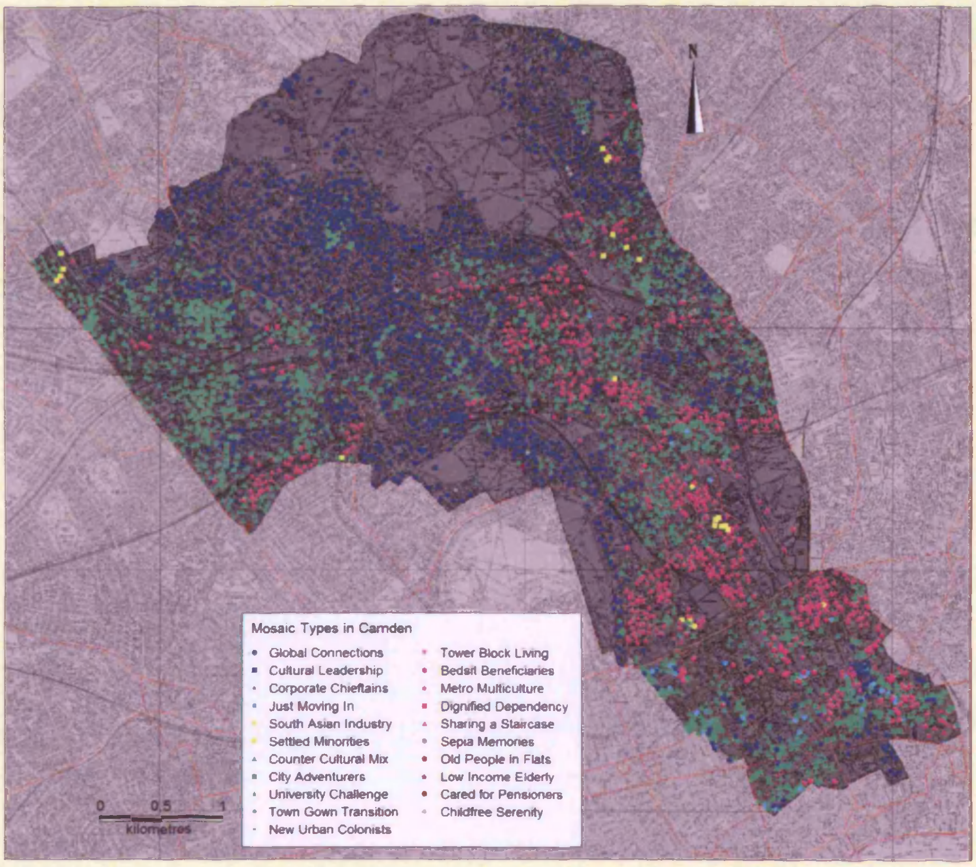


Figure 21: Map of neighbourhood Types in Camden

5.7 Chapter summary

This research method highlights the various data sources enumerated in section 5.3 which present themselves to the formation of a local (neighbourhood) spatial data infrastructure to form the building blocks for the development of health outcome profiles. Because the indicators being developed are potentially very useful to health professionals, with social marketing just one of the diverse applications for their use, it is pertinent to ground the research in practice and explore their dissemination with the public health sector. The following chapters, mark the beginning of section II, and document the development of the indicators and their application to explore the complexity of a number of different public health scenarios. Investigating health outcomes from two perspectives; diseases of comfort associated with lifestyle behaviours and as responses for improving disease prevention initiatives for various organisational scales. The results of the case studies are then synthesised together to consider their relevance as neighbourhood discriminators for social marketing, as alternatives to deprivation measures and as a method for combining and viewing social and geographical space.

Section II

Chapter 6

Geodemographic and health outcome analysis

Developing measures of diseases of comfort

6 Geodemographic and health outcome analysis – diseases of comfort

Objective One

> **To develop and explore an alternative framework for measuring and disseminating health outcome indicators of inequality, to be used in social marketing and public health interventions.**

6.1 Introduction

The literature review identified a current need for a population centric conceptualisation of measurement for defining health outcomes based on the social and geographical scale of the neighbourhood. This new framework should prove useful for a number of health applications that were highlighted in (Figure 13). The thesis itself concentrates on the particular domains of social marketing and health intervention planning. This is because they both have real world context and are of particular importance from the UK policy perspective, as highlighted in the literature review.

The four key principles of social marketing that have been adopted from the field of commercial marketing are; product, price, place and promotion, noted in section 3.4.2 of the literature review. With this in mind the forthcoming chapters will consider how these principles can be informed by geodemographics to assist in the development of interventions and campaigns. In the following chapters the research considers geodemographic classifications as a potential mechanism for appraising collective lifestyles, behaviours and health outcomes of population sub-groups with a view to understanding the four principles of social marketing.

Extensive research exists around defining and measuring social inequalities, with the underlying aim of government and health policy to reduce the inequalities that exist across social groups (Harper and Lynch, 2006). For these reasons, in the UK, there have been many discussions surrounding the reduction of health inequalities, as identified in the literature review, with the central aim of many key government targets leading to this common goal. The review in the literature proved that research into inequalities in health outcomes manifest through a variety of different social determinants as represented by the models identified in Chapter 2.3.2. One of the models is associated with the lifestyles of communities and individuals, and the population centric model by Friedman and colleagues (2004) noted the significance of collective responsibility. Furthermore, the current government emphasises the importance on both communities and lifestyles in its white paper: 'Choosing health' (DoH, 2004).

Therefore, in line with observations made by Marmot and colleagues (2005), this research will determine whether geodemographic indictors are able to identify and account for gradients in health inequalities which will be measured as the likelihood to partake in health-harming behaviours or develop a disease of comfort. This will be done by incorporating social similarity into the measurement technique by identifying socio-economic characteristics of small neighbourhood areas and ascribing the health-harming behaviours to these areas. Consideration will then be made to determine how the resultant indicators can be applied to real world public health issues. A series of case studies are used to develop the framework based on the policy priorities of public health departments. The results of the case studies are appraised and evaluated within the context of each case study and each one builds upon the knowledge gained from the previous study.

The production of useful robust health measurement indicators developed in the forthcoming chapters, are dependent upon data availability and the scale at which they are collected and made available. Health sector data use two primary collection techniques; transactional collection or quantitative surveying, qualitative data collection methods are used, but are predominantly used for small local areas (due to the limitations of cost both for time and resources). Transactional collection of health data involves recording transactions such as hospital visits, and patient lists maintained by general practitioners. Qualitative surveying techniques often involve face to face collection of data, between interviewer and respondent, which are subsequently recorded and coded for further analysis, but to maintain anonymity of respondents, data are released for large units of analysis e.g. postcode sector (Table 7).

The development of these new measures is important because, continuous demand is placed upon local PCTs to provide equitable services that ensure access for those who need them (Darzi, 2007, page 8). This represents a fundamental paradigm shift, moving away from a traditional ethos of universal health care for all, to one which places emphasis on a rationed system apportioned by need. The primary goals of health inequality initiatives are explicitly related to comparisons of outcomes between population groups and reducing known disparities. But there is an ever increasing pressure to improve the efficiency of service provision, whilst maintaining services that effectively serve the health needs of the population. It is essential to understand that different neighbourhoods and communities require different health promotion and screening strategies, and that different general practices will need to provide individual or shared services according to the profile of their patient lists, in order to maximise the benefits to all. With this in mind the case study approach should prove useful because it should enable successful differentiation of the population according to a diverse set of health outcomes.

In this chapter, Chapter 6, the annual Health Survey for England (HSE) is classified using a geodemographic classification to create contextual and compositional measures of neighbourhood lifestyles. This chapter explores the relevance and applicability of national survey data to create indicators for predicting lifestyle and behaviour patterns responsible for diseases of comfort such as obesity, smoking and drinking. The following chapter, Chapter 7 considers appropriate use of transactional health data for preventative health outcome measurements related to existing public health services, using Hospital Episode Statistics. The chapter will corroborate the geodemographics framework proposed in Chapter 4 and developed in Chapter 6, to predict diabetes need and uptake of cancer screening services together with the development of practice profiles that can be applied to assist understanding of the distribution of health need. In this sizable chapter the notion of scalable indicators, as required by the differing levels of decision making, is addressed by exploring the development of diabetes risk profiles for different organisational scales within the NHS such as general practices and Primary Care Trusts (PCTs). The chapter also considers the accuracy of the indicators by comparing results calculated using diabetes risk with a known and accepted industry model for predicting the population prevalence of diabetes.

6.2 Methodology and datasets

In this thesis a commercial geodemographic classification is used to create the risk index scores. The 2004 version of Mosaic (Experian, Nottingham, UK) was used in conjunction with a number of different health data sets (Table 8), in particular for this chapter it was used alongside the Health Survey for England (HSE), a description of which was outlined in Chapter 3.

In collaboration with Experian and with the assistance of the National Centre for Social Research (NATCEN), neighbourhood Types from the Mosaic geodemographic classification were appended to the annual national Health Survey for England (HSE) datasets, for the period 2001 to 2003 using the method described in Chapter 4.6, Figure 17. The neighbourhood Types were appended to the postcodes of the survey respondents. This process was carried out by NATCEN on their premises, in order to ensure compliance with data protection legislation. These data were then used to produce counts for each neighbourhood Type cross- tabulated with the responses to a range of HSE questions relevant to the behaviours described (see Table 10). A process flow diagram of the process in documented in Figure 22. In this diagram the division between processes carried out on site at NATCEN offices prior to data annonymisation are grouped together within the orange box. All other processes were carried out independent of NATCEN.

Once the geodemographic classification had been appended to the HSE for the years 2001, 2002, and 2003, the response variables were cross-classified with both geodemographic Group and Type. Of concern to this investigation were the responses to questions pertaining to smoking, obesity and drinking; all lifestyle behaviours that are often linked to diseases of comfort. The response questions considered to be of interest to this research are listed in Table 10.

Table 10 lists three types of attributed behaviours that are admissible. The first corresponds to the frequency of smoking behaviours and questions survey respondents about how many cigarettes they smoke per day. The second set of question responses is considered a proxy measure for obesity because it records the Body Mass Index (BMI) which is a height to weight ratio of the responders. The third and final set of responses relate to the alcoholic drinking patterns of responders by asking how many drinks per week you have.

Response Variable	Attribute level
Frequency of Smoking	Light smokers, under 10 a day
	Moderate smokers, 10 to under 20 a day
	Heavy smokers, 20 or more a day
	Don't know number smoked a day
	Non-smoker
	No answer/refused
Body Mass Index (BMI)	BMI Under 18.5
	BMI between 18.5 and less than 25
	BMI between 25 and below 30
	BMI between 30 and below 40
	BMI 40 and more
	BMI not applicable
Number of drinks had in last 7 days	no drink in last 7 days
	Less than 2
	2 to 3 drinks
	3 to 4 drinks
	4 to 5 drinks
	5 to 6 drinks
	6 to 8 drinks
	8 or more drinks

Table 10: Table of Health Survey for England health behaviour variables

The response variables and the corresponding attribute level data listed in Table 10 were extracted from the dataset, and all counts for males and females over the age of 16 (all adults), for each year by neighbourhood Type were aggregated together. In total the research database had 37,851 adult counts by Mosaic Type for the three years.

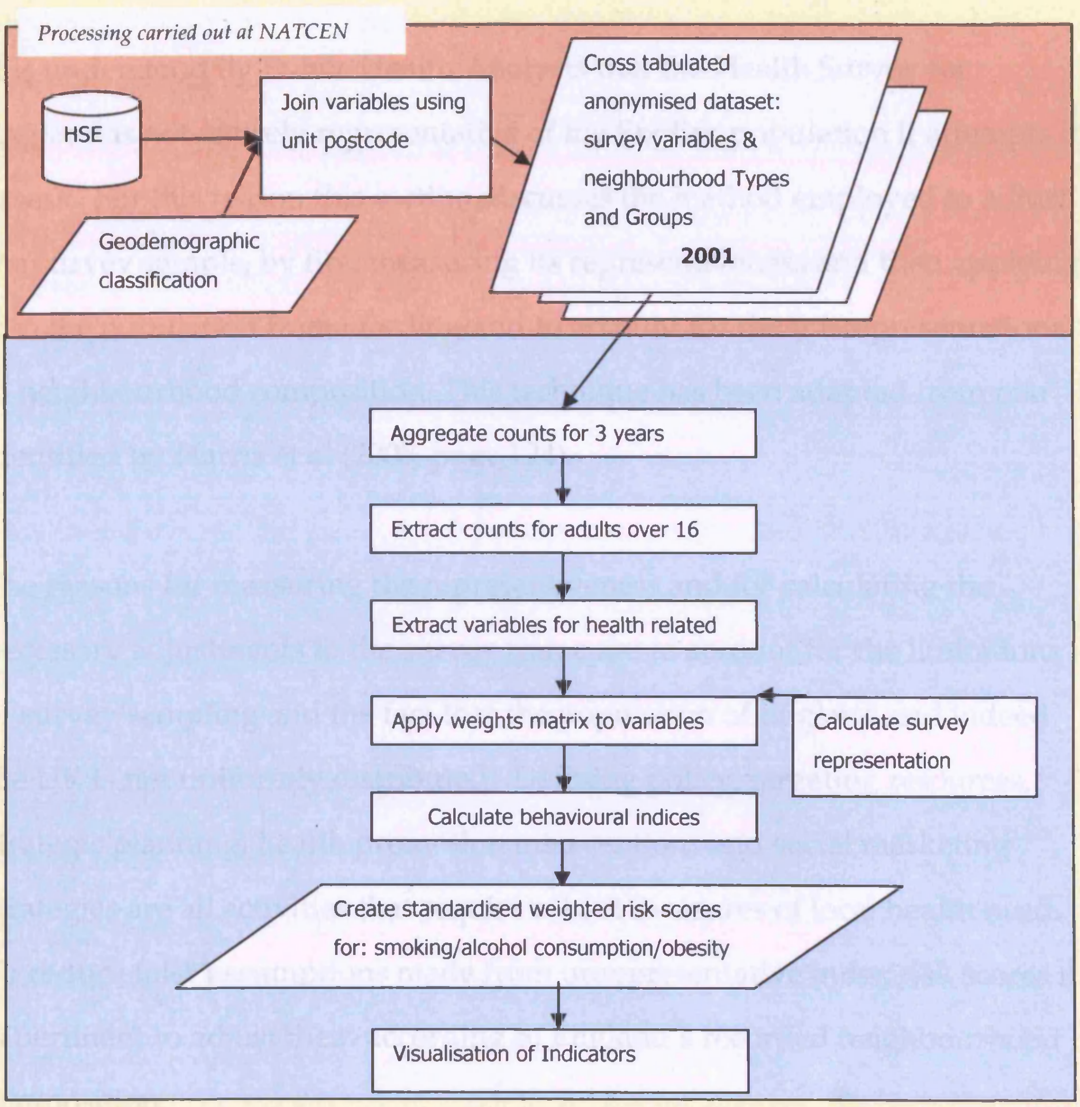


Figure 22: Process flow diagram of method used to create indices for measuring the health-harming behaviours of smoking, drinking and obesity

The next step in the process was to produce standardised weighted risk scores for each the response variables and their associated attribute levels listed in Table 10: smoking, drinking and obesity (calculated using BMI)). The standardised risk scores are weighted because for some neighbourhood

Types the HSE survey methods produced over-sampled results and for others it produced under sampled results. This is discussed in greater detail in the next section.

6.2.1 Representiveness of survey data

It is understood by Public Health Analysts that the Health Survey for England is not entirely representative of the English population it attempts to assess. For this reason this section discusses the method employed to adjust the survey sample, by first measuring its representiveness and then applying it to the population frame for England to account for the misrepresentations in neighbourhood composition. This technique has been adapted from one identified by Harris et al (2005, page 124).

The reasons for measuring the representiveness and for calculating the necessary adjustments to the survey frame are to account for the limitations of survey sampling and the fact that the population of England and indeed the UK is not uniformly distributed. Defining policy, targeting resources, strategic planning, health promotion interventions and social marketing strategies are all activities that require robust measures of local health need. To reduce false assumptions made from unrepresentative index risk scores it is pertinent to adjust them according to England's recorded neighbourhood composition.

The number of people in different population sub-groups varies, for example only 1.97% of the UK's population live in Neighbourhood Type 36 (see Table 11). Consequently, because of the different neighbourhood compositions in England there is a possibility that sampling methods may subject the dataset to being over or under representative. The under or over representation of neighbourhoods is an artefact of the sampling frame used to conduct the survey, where the population is first divided into geographical areas, but not population sub-groups, before it is sampled.

There are a number of steps to ensure the survey is adjusted appropriately. First the 'All Fields Postcode Directory' (AFPD) was extracted from UK Borders, (also known as Gridlink). This dataset contains a record of each postcode unit in the United Kingdom. It relates both current and terminated postcodes to a range of current statutory administrative, electoral, health and other area geographies (ONS Geography, 2005, page 4). All postcodes with a country code of 064, representing England were extracted. To standardise their format, the postcodes were stripped of their blank spaces to ensure they were all in the same format then linked to the Geodemographic classification which records the total number of households and population per postcode unit. From this the total population living in neighbourhood Types in England (see Table 11) was calculated, and forms the population base used to calculate the index scores in future calculations.

Second, the validity of the survey was established by calculating index scores of the HSE respondents for each of the variables. Compared to England's total population per response variable, this determines the extent to which the survey represents the population. The standard formula for calculating indices, listed in section 5.2 was used, but for this example the numerator was the percentage of total survey respondents per neighbourhood Type (for each response variable and the denominator the percentage of England's population per neighbourhood Type). For example using neighbourhood Type 01, the proportion of the total number of survey respondents (column 2, Table 11), was divided by the proportion of that Type in England (column 4, Table 11) [0.0053/0.0066=0.812 (to 2 decimal places)]. The result of this calculation returns an index score of the survey representation for each neighbourhood. In this instance Type 01 is under represented (result is less than 1) in the survey. The results for all neighbourhood Types are listed below in Table 11.

Mosaic Type Code	Total respondents to BMI survey variables	Proportion of total survey respondents (% target, to 4 decimal places)	Count of Type in England	Proportion of Type in England (% Base, to 4 decimal places)	Representation of survey variables (sampling index score, to 3 decimal places)	Weight to be applied to BMI index score calculations
01	202	0.0053	339,705	0.0066	**0.812****	1.232
02	391	0.0103	519,269	0.0101	**1.028***	0.973
03	506	0.0134	708,194	0.0137	**0.975**	1.025
04	688	0.0182	756,400	0.0146	**1.242**	0.805
05	729	0.0193	874,829	0.0169	**1.137**	0.879
06	892	0.0236	1,197,831	0.0232	**1.016**	0.984
07	964	0.0255	1,283,913	0.0249	**1.025**	0.976
08	55	0.0015	240,937	0.0047	**0.312**	3.209
09	454	0.0120	657,300	0.0127	**0.943**	1.061
10	558	0.0147	872,459	0.0169	**0.873**	1.145
11	964	0.0255	1,375,963	0.0266	**0.956**	1.046
12	1291	0.0341	1,713,993	0.0332	**1.028**	0.973
13	758	0.0200	1,103,341	0.0214	**0.938**	1.066
14	55	0.0015	112,936	0.0022	**0.665**	1.504
15	1383	0.0365	1,718,624	0.0333	**1.098**	0.910
16	1365	0.0361	1,562,226	0.0302	**1.193**	0.838
17	1251	0.0331	1,653,580	0.0320	**1.033**	0.968
18	1433	0.0379	1,930,795	0.0374	**1.013**	0.987
19	995	0.0263	1,449,480	0.0281	**0.937**	1.067
20	428	0.0113	771,361	0.0149	**0.757**	1.320
21	1060	0.0280	1,365,134	0.0264	**1.060**	0.944
22	1345	0.0355	1,691,541	0.0327	**1.085**	0.921
23	1589	0.0420	2,129,553	0.0412	**1.018**	0.982
24	1069	0.0282	1,433,884	0.0278	**1.018**	0.983
25	307	0.0081	436,928	0.0085	**0.959**	1.043
26	418	0.0110	719,960	0.0139	**0.792**	1.262
27	773	0.0204	977,535	0.0189	**1.079**	0.926
28	407	0.0108	686,775	0.0133	**0.809**	1.236
29	295	0.0078	521,115	0.0101	**0.773**	1.294
30	525	0.0139	718,196	0.0139	**0.998**	1.002
31	479	0.0127	582,131	0.0113	**1.123**	0.890
32	285	0.0075	447,299	0.0087	**0.870**	1.150
33	238	0.0063	358,767	0.0069	**0.905**	1.104
34	106	0.0028	251,753	0.0049	**0.575**	1.740
35	109	0.0029	144,025	0.0028	**1.033**	0.968
36	531	0.0140	994,608	0.0193	**0.729**	1.372
37	459	0.0121	736,541	0.0143	**0.851**	1.176
38	109	0.0029	120,283	0.0023	**1.237**	0.808
39	273	0.0072	358,337	0.0069	**1.040**	0.962
40	7	0.0002	10,023	0.0002	**0.953**	1.049
41	436	0.0115	744,068	0.0144	**0.800**	1.250
42	867	0.0229	1,378,362	0.0267	**0.859**	1.165
43	965	0.0255	1,206,565	0.0234	**1.092**	0.916
44	1046	0.0276	1,254,419	0.0243	**1.138**	0.879
45	812	0.0215	1,027,439	0.0199	**1.079**	0.927
46	1269	0.0335	1,800,956	0.0349	**0.962**	1.040
47	980	0.0259	1,425,527	0.0276	**0.938**	1.066
48	176	0.0046	221,098	0.0043	**1.087**	0.920
49	493	0.0130	587,870	0.0114	**1.145**	0.874
50	422	0.0111	524,897	0.0102	**1.097**	0.911
51	236	0.0062	265,965	0.0051	**1.211**	0.826
52	468	0.0124	561,083	0.0109	**1.139**	0.878
53	629	0.0166	742,943	0.0144	**1.156**	0.865
54	504	0.0133	595,835	0.0115	**1.155**	0.866
55	976	0.0258	1,261,151	0.0244	**1.056**	0.947
56	70	0.0018	108,136	0.0021	**0.884**	1.132
57	95	0.0025	75,498	0.0015	**1.718**	0.582
58	638	0.0169	1,002,648	0.0194	**0.869**	1.151
59	538	0.0142	702,187	0.0136	**1.046**	0.956
60	377	0.0100	521,948	0.0101	**0.986**	1.014
61	108	0.0029	128,632	0.0025	**1.146**	0.873

Table 11: Weightings of Health Survey for England response variables of Body Mass Index

*Values in red indicate over-sampling of responses, ** values in blue in under-sampling of responses

The resultant index score for each neighbourhood Type signifies how representative the survey was according to the total number of people who responded to the particular survey question. Neighbourhood Types with scores less than 1 were indicative of under-sampling in neighbourhoods. With scores greater than 1 suggesting the neighbourhood Types were over-sampled. In the case of the worked example, responses to questions related to BMI showed 28 neighbourhoods were under-sampled (coloured in blue in Table 11). Some neighbourhoods were over represented indicating over-sampling of the survey, this was evident in 33 of the 61 neighbourhoods (coloured in red in Table 11).

6.2.2 Weighting the Health Survey for England data to account for sampling errors

Returning to Figure 22, the next step in the process was to calculate weighted index scores that accounted for representation errors in the sampling frame of the survey. The resultant scores for the survey representation were used to assist in the creation of the sampling weights. Each index score is calculated using a scale where 100 represents the national average. The weights were defined in the following way and the results of which are included in column 7, Table 11. For Neighbourhood Type 01, the weight would be [100/ (0.812*100) = 1.232]. So first the proportion of over/ under-sampling per neighbourhood was identified. Neighbourhood Type 01 was under sampled by 23%. The population counts must be adjusted by +23%, hence a weight of 1.23.

6.2.3 Adjusting the survey counts to calculate behavioural and lifestyle indicators

The final steps in the process flow diagram (Table 11) were associated with calculating the final weighted survey variables to return health outcome risk indicator scores. By applying the weights calculated in the previous section it is possible to adjust the survey counts according to whether the neighbourhood Type was over or under sampled in relation to the population distribution of neighbourhood Types in England. The process once again follows that suggested by Harris et al (2005, page 125). If the proportion of survey respondents compared to the proportion of the actual population in England was under-represented (i.e. scores are under 1) then they are given a higher weight, and those over 1 are given a lower weight. A series of calculations were used to do this, but in essence the final index score is produced in the same way as that noted in Chapter 5, section 5.5.

The actual count of the survey respondents were adjusted according to the weights calculated in the previous section. So returning once again to the example of neighbourhood Type 01, the sampling index score showed that for questions related to BMI, this neighbourhood Type was under represented in the survey by a factor of 0.812 (23%). The weight to be applied to this Type was calculated as a 1.232, the factor which must be used to adjust the response counts for Type 01. This is done by taking the original number of survey respondents and multiplying it by this weighting factor (202*1.232 = 249). Therefore the adjusted number of respondents used for risk scores associated with BMI now becomes 249, 47 more people were added to this response count. These calculations are repeated for each neighbourhood Type and become the new input variables for the nominator of equation 1, Chapter 5, section 5.5.

6.2.4 Summary

It is on this weight adjusted dataset that the case studies in this chapter have primarily focused, investigating 3 lifestyle behaviours: smoking, obesity and drinking. The complex nature of these behaviours and absence of routine data collection reinforces the difficulty of exploring these lifestyles at the unit of the neighbourhood. The case studies presented in this chapter develop the framework for applying geodemographics to national surveys, exploration of resultant neighbourhood patterns and discusses the relevance for application to social marketing and as an alternative predictor for health outcomes. Where the method goes beyond the development of a simple indicator as discussed in the above methodology more detail is discussed within the context of each case study.

6.3 Predicting smoking behaviours

6.3.1 Problem definition

Smoking related illness impact significantly on the NHS and its services. In 2004 the Department of Health measured smoking prevalence at 25% for adults living in England (Information Centre, 2006, page 3). The same source attributes 1.4 million hospital admissions with a primary diagnosis directly related to smoking.

Smoking behaviour is difficult to measure and predict at neighbourhood scales, due to the scarcity of available data. This is because there is no central repository of data on smokers or smoking related behaviour in the UK. Data are sometimes managed by GPs, but these data are rarely up to date and aside from prevalence rates for general practices there is little data sharing between general practices and PCTs. Currently recording of smoking prevalence is not mandatory for general practices, nor is sharing such data with PCTs obligatory.

Despite this lack of data, a key national target upon which PCT performance is measured relates to the success of smoking cessation services. For this reason there is a heavy onus on PCTs to reduce the number of smokers by increasing the number of people who successfully quit the habit. The problem arises of how to identify population sub-groups that are most likely to comprise of smokers and then encourage them to quit.

In this case study a number of different applications of geodemographics are employed to help facilitate the exploration of the social and spatial distribution of smokers. In the first example the Health Survey for England (HSE) is cross-referenced with the neighbourhood classification to create cumulative proportions of smokers per neighbourhood in order to investigate the evenness of their distribution in society. This is done using Lorenz curves and subsequently calculating the associated Gini coefficients. The motivation for calculating these measures is to carry out initial inductive analysis in order to determine whether smoking behaviour is distributed disproportionately in society. The second application of the geodemographic framework carries out exploratory analysis by creating weighted risk index scores (as discussed in section 6.2) to predict neighbourhood propensity to smoke or not to smoke. Furthermore, these risk scores are then converted into probabilistic density surfaces for London which are then used in a following case study to create a composite surface of unhealthy behaviours.

6.3.2 Exploratory analysis using Lorenz curves and Gini coefficients

The measurement of health inequalities has long been researched and the link between smoking behaviours and poor health outcomes is well known and irrefutable. But despite the common knowledge surrounding the dangers related to smoking, one quarter of the adult population in the UK partake in this negative health behaviour. In the 2004 Wanless report smoking was identified as,

> *"the single greatest cause of preventable illness and premature death in the UK. …it is estimated that half the difference in survival to 70 years of age between social class I and V is due to higher smoking prevalence in class V"* (Chapter 4, page 2).

Currently the pattern between smoking and income is measured by social class or using a deprivation measure (ASH, 2005).

In health research a common technique for assessing disproportionalities include; the Gini-index, the squared coefficient of variation and the relative concentration index; for a more comprehensive list see Harper and Lynch (2006). To begin exploring the variation in smoking related behaviours for neighbourhoods, initial analysis took the commonly used technique to explore income inequalities known as Lorenz curves and applied it to the smoking dataset.

Initial descriptive analysis involved the creation of Lorenz curves and Gini coefficients to explore the equity of the distributions. As the name suggests Lorenz curves were developed by Lorenz, in 1905, and have been frequently used to determine the proportion of income in relation to the percentage of the population. In this example the technique is applied to determine the proportions of differing health behaviours as percentages of the survey sample to summarise differences in health behaviours between different social groups.

The raw counts of variables cross-referenced by geodemographic Type were used as the input data (Table 9). The variables representing people who were classed as moderate and heavy smokers were summed together, (light smokers were not included because initial examination of the data distributions highlighted a different pattern to that of moderate and heavy smokers and thus are treated as a separate group). This created a new variable labelled 'smokers (SM)' and was compared to the people who were defined as 'non smokers (NS)'. These variables were used to create a population ratio of moderate to heavy smokers compared to non-smokers (SM/NS) for each neighbourhood Type. The neighbourhoods were then ranked, in ascending order, based on these ratios. Then for each neighbourhood the total numbers of smokers (% SM) or non-smokers (% NS) was converted into percentages. These values were accumulated in accordance with the previously calculated ascending ranks and plotted against each other on a graph.

The graph in Figure 23 plots the percentage of non-smokers versus the percentage of smokers for each given neighbourhood Type (using the Mosaic geodemographic classification). Each point on the graph represents one of the 61 different Neighbourhood Types in the geodemographic classification. The Lorenz curve shown indicates the degree of difference across neighbourhood Types for smoking related behaviours. If smoking behaviours across neighbourhoods were perfectly equal, then the distributions would be proportionally equal and the plot would be a 45 degree diagonally straight line (Kawachi and Kennedy, 1997, page 1123). More unequal distributions are highlighted by the greater concavity to the shape of curve. By investigating the shape of the curve the amount of inequality can be ascertained.

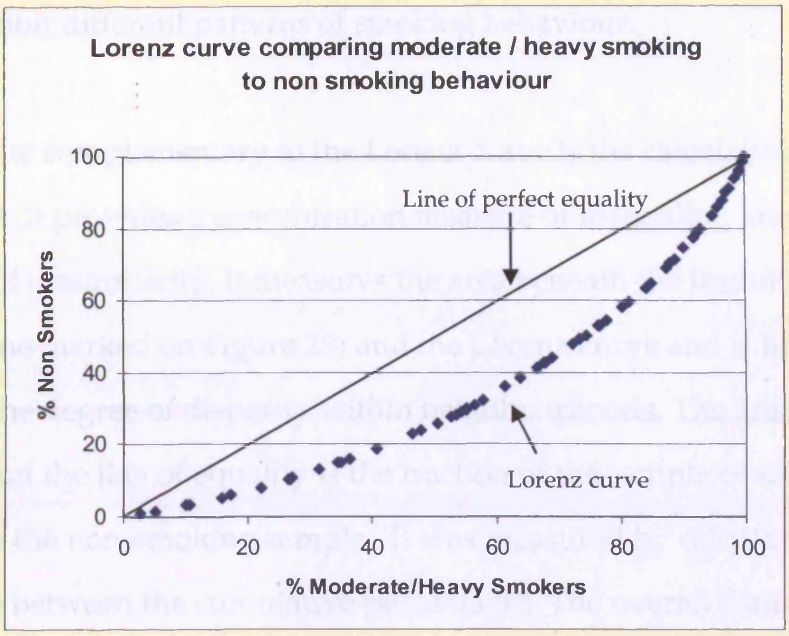


Figure 23: Lorenz curves for smoking patterns across neighbourhood Types

If smoking related behaviours vary at the neighbourhood level, then consideration should be made for the appropriateness of interventions planned at this scale and indicates the relevance for further analysis.
The measure in the context was used to calculate disproportionate amounts of smoking in neighbourhoods. Moderate and heavy smoking behaviours can be considered as an association between neighbourhood's share of the population, ranked by their non-smoking behaviour and their share of non-smoking behaviour.

The graph in Figure 23 highlights the degree of inequality in the phenomenon being measured, in this instance smoking behaviours within neighbourhood Types. The concavity of the curve produced by comparing the distributions of smokers compared to non-smokers within neighbourhoods is pronounced. The chart indicates a degree of difference between the distributions, implying that at the neighbourhood level in England there is a measure of separation between this type of smoking behaviour. The results indicated that for neighbourhood Types in England there is degree of difference between proportions of moderate/ heavy smokers and non-smokers. Neighbourhood Types exhibit different patterns of smoking behaviour.

A technique complementary to the Lorenz curve is the calculation of the Gini coefficient. It provides a concentration measure of inequality, known as the measure of dissimilarity. It measures the area beneath the line of equality (the straight line marked on Figure 23) and the Lorenz curve and is used to describe the degree of disparity within neighbourhoods. The area between the plot and the line of equality is the fraction of the sample of smokers relative to the non-smoking sample. It was measured by calculating the difference between the cumulative percentages. The overall Gini coefficient for English neighbourhoods was measured as 25.41(+-1.97) and has a range of 0 to 100, where 0 represents total equality and 100 represents total inequality. It highlights the degree to which the distributions depart from equality. Thus, smoking distributions in neighbourhoods across England can be said to be unequal. Neighbourhoods with high proportions of moderate to heavy/smokers appear to have concentrations of that behaviour and are less likely to have concentrations of non-smokers.

Gini coefficients for each neighbourhood Type were also calculated enabling the calculation of the degree of disparity within neighbourhood Types. These values were mapped using a Geographical Information System (GIS), whereby each postcode in the UK has an associated neighbourhood Type and each unit postcode area has a corresponding set of spatial coordinates. The Gini coefficients were adjoined to spatial coordinates according to their neighbourhood Type and mapped using MapInfo. The neighbourhoods with the largest Gini coefficient represented the within variation of smoking related behaviour. The neighbourhood Type with the largest within variation of coefficients was Type 36; a neighbourhood representing diverse communities often found in inner city areas with people most likely to be living in public housing and struggling financially. This indicates that social marketing campaigns targeting Type 36 would most likely to be comprised of large numbers of smokers, and thus worthy of targeting, and are highlighted in Figure 24.

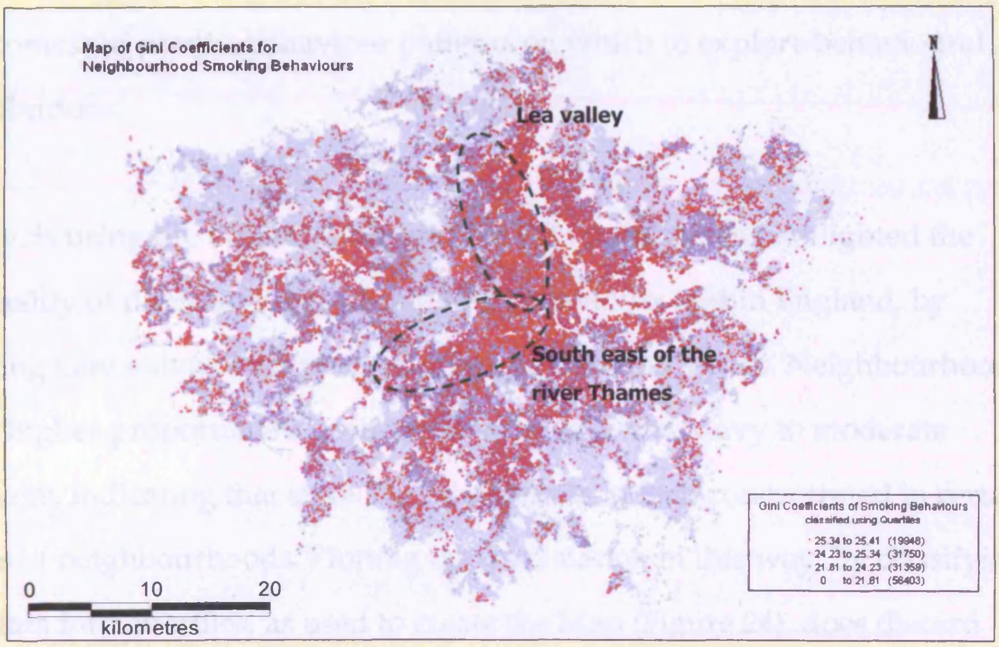


Figure 24: Spatial distribution of Gini coefficients for smoking related behaviour

Figure 24, highlights the distribution of Gini coefficients for smoking related behaviours in London. Each point on the map represents the centroid of a neighbourhood postcode unit. Neighbourhoods coloured in dark red and purple have the highest individual Gini coefficients, representative of locations where the within differences between smoking behaviours were likely to be greatest. Neighbourhoods with greater tendencies to smoke and have higher concentrations of smokers were more likely to be found on the south east bank of the river Thames and on the west side of the Lee valley. Once again from a public health view, social marketing initiatives could be effective in these areas and are worth targeting.

6.3.3 Exploratory analysis using geodemographic indices

The inductive analysis using the Lorenz curve and Gini coefficient indicated that smoking related behaviour varied at the level of the neighbourhood, and using this survey response variable would be useful for further exploration. The next step in the analysis was to create small area measures of risk (outcomes) to predict behaviour patterns on which to explore behavioural distributions.

Analysis using the Lorenz curve and the Gini coefficients highlighted the inequality of different types of smoking behaviours within England, by plotting Gini values associated with neighbourhood Types. Neighbourhoods with higher proportions of non-smokers have fewer heavy to moderate smokers, indicating that smoking behaviour is indeed concentrated in certain types of neighbourhoods. Plotting the information in this way, by classifying the data into quartiles, as used to create the Map (Figure 24), does discard some information but can be justified because of its usefulness as a communication device. Further detailed exploration of the dataset and relationships of smoking across neighbourhoods is conducted in the following section.

The production method of standardised neighbourhood indices is similar to location quotients, which have been widely used throughout economic geography and location analysis and provide the facility to compare the counts of a local activity relative to a regional or national base. In this instance their use is similar but matches sample estimates of population characteristics and attitudes with local conditions and does not account for local activity within a unit, but neighbourhood activity according to socio-economic similarity. Section 5.2 of the research framework outlined the method used to create the neighbourhood indices.

At this juncture it is worth noting that the soon to be calculated measures do not compare actual local incidence values, but calculate estimated risk rates according to ascribed population characteristics. The standardised and weighted neighbourhood index scores produced in this research facilitate the proxy measurement of the variability of health behaviours across different population sub- groups, and can be compared to a national baseline for England which in this instance will be given a value of 100.

6.3.3.1 *Weight adjusted smoking risk*

This section describes how the geodemographic framework is extended to calculate adjusted risk index scores for predicting smoking behaviour at the scale of the local neighbourhood. To do this, weights were first calculated for each of the smoking response variables from the HSE (Table 12) following the method laid out is section 6.2. The survey counts for each neighbourhood were adjusted in relation to their ability to reflect the population of England. For each smoking related variable, see Table 12, the counts of respondents within each neighbourhood Group were calculated. The variable counts were then weighted, using the method from the previous case study. The weighted counts were then substituted into the standard index equation (section 5.3). In this instance the numerator equates to the weighted counts per variable whilst the denominator remains as the percentage of England's population. To reiterate, if calculating a measure to predict the neighbourhoods who have a likelihood to be light–smokers, the first step is to determine if the respondents for that question in each neighbourhood reflect the population of England or whether they are over/under-represented. This is done by calculating the proportion of respondents per neighbourhood, compared to the proportion of England population per Type, so for Type 01: [(202/ 37,666) / (339,705/51,951,846) = 0.82]. This value indicates that for smoking questions Type 01 is under-represented in the population. The resultant weight is then calculated for that neighbourhood Type ((100/0.82*100)) the result means that original survey counts must be adjusted by a factor of 1.22, reflecting the weight to be applied. The light smokers index score is then calculated using this weight applied to respondent counts per neighbourhood for those who were classified as light smokers, [((27*1.22)/2806) / (339,705/51,951,846)=180].

ID	Variable Description
1	Light smokers, under 10 a day
2	Moderate smokers, 10 to under 20 a day
3	Heavy smokers, 20 or more a day
4	Don't know number smoked a day
5	Non-smoker
6	No answer/refused
7	Don't know
8	Item not applicable

Table 12: Smoking variables extracted from Health Survey for England (years 2001,2002, 2003)

This calculation was then applied to all response counts per variable in Table 12, which outlines the smoking variables on the HSE. With the exception of respondents who did not answer the question, replied "don't know" or for whom the question was not applicable, these counts were included in the initial calculations for determining the weights, but not to create the final risk scores of smoking behaviour because their habits were not known.

The results showed variation in neighbourhood patterns of smoking related behaviours. Table 13 shows the minimum and maximum risk scores calculated and the purpose of their inclusion is to give an indication of the breadth of difference in smoking behaviour across the neighbourhoods. The patterns follow those of the previously calculated Lorenz curve and Gini coefficient. The biggest difference is seen for the variable representing moderate smoking behaviours, where 100 represents the average in England. Neighbourhoods above 100 are more likely to be composed of moderate smokers and neighbourhoods below this threshold are unlikely to be moderate smokers. Neighbourhood Type 40 (Sharing a Staircase) are 16 times more likely than Type 03 (Corporate Chieftains) to smoke a moderate number of cigarettes. People living in neighbourhoods representative of families on benefits have the highest likelihood of being heavy smokers and are the least likely not to smoke at all. Indeed they are more than twice as likely to be heavy smokers as compared to the national average, and the likelihood that they are heavy smokers is 10 times greater than the neighbourhoods in which residents are least likely to smoke. At this junction it is necessary to point out that there one limitation with analysis of this type. The categories chosen by the HSE to group smoking attributes of responders will, to some extent, condition the results of this analysis and all types of analysis using this dataset.

	All Smokers Index (100 = national average)	**Non Smokers Index**	**light smokers Index**	**Heavy Smokers Index**	**Moderate Smokers Index**
Min	35	66	38	24	18
Max	205	124	227	251	285
St dev	42.91	15.17	43.50	56.94	56.20

Table 13: Simple statistics for describing the distribution of index scores of smoking attributes across neighbourhood Types

In order to measure non-smoking behavioural patterns, 100 was again used to represent the baseline England average, but this time neighbourhoods with index scores greater than 100 are indicative of communities who are least likely to smoke. The results indicated that non-smokers are most likely to be found in neighbourhoods of affluent older families representing Type 04 (Golden Empty nesters). The standard deviations of the index scores are also recorded in Table 13. The heavy smoker's index has the largest standard deviation (56.2), indicating that this behaviour is much more widely dispersed across neighbourhoods. The distribution for non-smokers has a considerably lower standard deviation (15.2).

Figure 25 highlights the weighted index scores for the predominant neighbourhood Types within the Inner London Borough of Camden, the main test bed for this research. The five smoking related variables and their associated risk index scores have been displayed on the graph. The benchmark of 100 was once again used as the average for neighbourhoods in England. The most affluent neighbourhoods, Type 01 (Global Connections, highlighted in dark purple), are the most likely not to smoke, but at the same time are also likely to be a residential population of light smokers. These light smokers are a potentially useful target group, as smoking is not ingrained within their social Type and moreover they are not likely to have a deep-seated addiction so may find giving up easier than other Types.

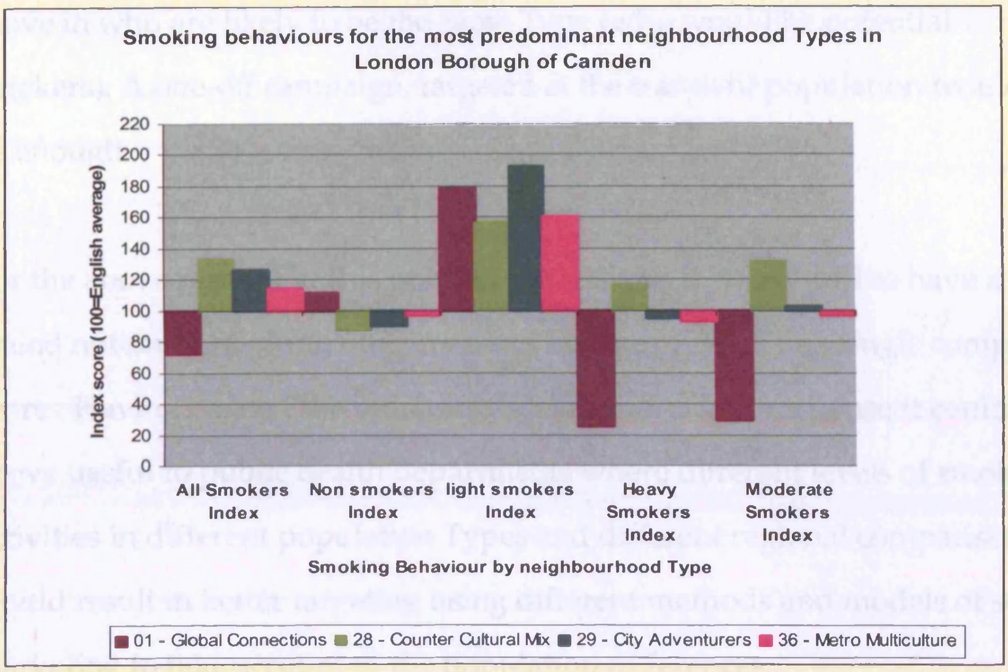


Figure 25: Smoking behaviours for most frequently occurring neighbourhood Types in Camden, London.

Staying with Camden for the meantime, the variable created for all smoking related behaviours (all smokers index) in Figure 25 shows that three out of four neighbourhood Types are likely to be smokers as compared to the national average. Young transient singles living in rented accommodation (Type 28 – in light green) are most likely to smoke and have a range of smoking habits; light, moderate or heavy smokers.

From both a policy and social marketing perspective this is a potentially useful finding. Many initiatives are targeted at low-income and deprived neighbourhoods representing low socio-economic classes commonly associated to Type 36, and not at the young transient populations represented by Type 28. The results of these profiles, indicates a need for more than one initiative for the different population sub-groups. Bearing in mind the transient nature of Type 28 consideration should also be made in budgeting and planning of marketing campaigns and interventions. They would need to be synchronised with the transient cycles of the population as new people move in who are likely to be the same Type (who would be potential smokers). A one-off campaign, targeted at the transient population would not be enough.

For the above reasons at this point in the analysis it was useful to have a scaled notion of smoking outcomes and not simply look at a single composite score. Having created the indictors for each individual response it could prove useful to public health departments where different levels of smoking activities in different population Types and different regional comparisons would result in better targeting using different methods and models of social marketing to take account of the population differences.

Exploration of the patterns of the index scores enabled the following null hypothesis to be developed: A significant pattern does not exist between neighborhood Type and the different response variables used predicting smoking related risk behaviour.

Both the Gini coefficients and the Lorenz curves from the previous section highlighted how smoking attributes vary amongst different populations sub-groups. An alternative to using these measures is to plot the derived neighbourhood index scores for each of the smoking attributes against each other. This takes the form of a correlation matrix. The results enable the examination of the patterns between different types of smoking related behaviours within the different neighbourhoods in England. The purpose of which is to indicate whether the calculated risk index scores for different smoking attributes are unique to that neighbourhood or if neighbourhoods are comprised of different types of smokers. For social marketing this matrix is important because it will aid understanding of the types of interventions that need to be created. Neighbourhoods comprised of a mix of smoking behaviours will require unique stop smoking strategies for each different attribute. For example, those neighbourhoods that have both a high index scores for light smoking and heavy smoking would require distinct strategies for enabling the heavily addicted to change their deep-seated behaviours to those who only smoke a few a day with an milder addiction.

The correlation matrix enabled the description of the different relationships between each of the 5 smoking response variables and how they vary with each other, all smokers, non-smokers, light smokers, moderate smokers and heavy smokers. The index scores for one smoking attribute was plotted against the index scores for all the remaining variables. The graphical representation of the resultant correlation matrix is highlighted in Figure 26. The graph is divided into 25 grid squares each representing one bivariate plot. The bottom left corner of each grid represent low index scores (0) and as you progress along each individual grid as the x and y axis increase so do the values of the risk scores. It shows the patterns of association for the different smoking variables across neighbourhood Types. On the chart each circle represents one of the 61 neighbourhood Types.

On first inspection of Figure 26, there appears to be a number of different associations at play. For example, neighbourhoods with high index scores for non-smoking also have low scores for the composite smoking index (Figure 26, label a), which is in line with the Gini coefficients examined earlier in this chapter.

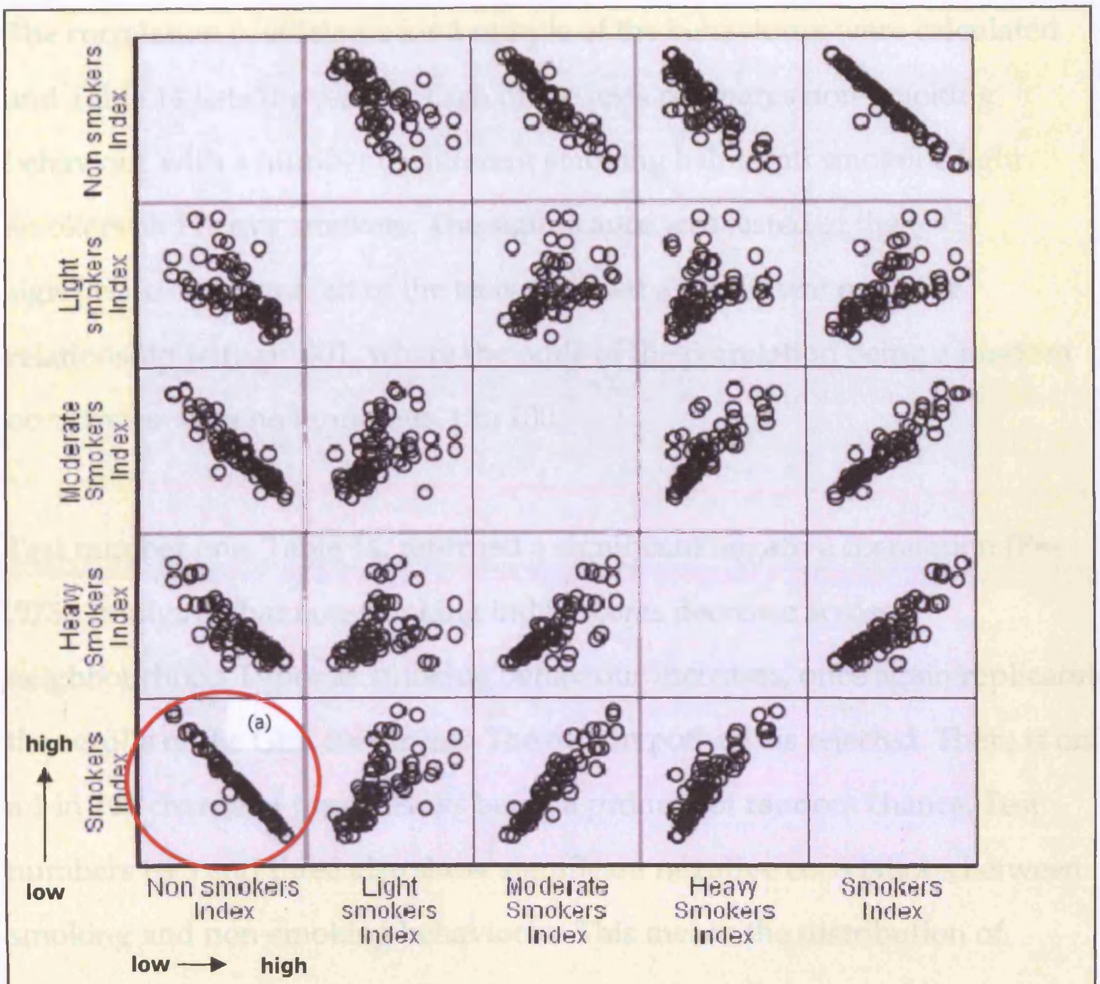


Figure 26: Correlation matrix for risk index scores of smoking related behaviour across neighbourhoods

To determine if the relationships had any significance they were explored further using Spearman's rank correlation. The equation used to calculate Spearman's rank is noted below. Spearman's rank correlation coefficient is denoted by r_s and the D represents the calculated differences between the ranks of the scores for the different smoking variables.

$$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)} \qquad \text{Equation 2}$$

The correlation coefficients for a sample of the behaviours were calculated and Table 14 lists the results. Each of the tests compares non-smoking behaviour with a number of different smoking habits; all smokers, light smokers and heavy smokers. The significance was tested at the 0.01 significance level and all of the tests returned a significant negative relationship with $p<.001$, where the odds of the correlation being a random occurrence were no more than 1 in 100.

Test number one, Table 14, returned a significant negative correlation ($P=-.975$), implying that non-smoking index scores decrease across neighbourhood Types as smoking behaviour increases, once again replicating the results of the Gini coefficient. The null hypothesis is rejected. There is only a 1 in 100 chance of these results being a product of random chance. Test numbers two and three also show significant negative correlations between smoking and non-smoking behaviours. This means the distribution of different Types of smoking related behaviour for different neighbourhood Types is significant.

Test number 1 – non-smokers vs smokers			Smokers Index
Spearman's rank	Non smokers index	Correlation coefficient	-.975
Test number 2 – non smokers vs light smokers			**Light smokers Index**
Spearman's rank	Non-smokers index	Correlation coefficient	-.732
Test number 3 - non smokers vs heavy smokers			**Heavy Smokers Index**
Spearman's rank	Non-smokers index	Correlation coefficient	-.831

Table 14: Spearman's rank correlation coefficients for a sample of smoking behaviours

The results indicated that targeting neighbourhoods using different marketing techniques for different neighbourhoods would have significant value. This is because there was significant variation of distinctive smoking attributes of comprising residential populations within individual neighbourhood Types. Targeting these different neighbourhoods according to their predominant smoking attributes together with the most appropriate behavioural change strategy for the type of smoking, would most likely, improve accuracy and efficiency of campaigns and lead to a reduction in the number of smokers. This is because the patterns of predicted risk to smoking behaviour are significantly different within the same neighbourhoods and across different neighbourhoods

6.3.4 Mapping smoking risk using kernel density estimation

To investigate the spatial distribution of the risk scores, neighbourhood scores associated with smoking behaviours were then mapped using point density estimation in a GIS. This method was chosen because it is commonly used within public sector organisations to create hotspot maps; but it is not strictly a methodology for defining spatial clusters (de Smith et al, 2008 online). This is because the technique facilitates the development of the spatial representations of population health at small levels, through the production of density surfaces. As acknowledged by Lloyd (2005) there has been a move towards the use of surfaces to re-represent the original points in a more useful and visually understandable format (Martin, 1996 and Martin **et al.**, 2000). The purpose of doing so in this research is to provide a generalised pattern of smoking behaviour and to indicate likely spatial clusters.

The process used for creating surfaces in this research is known as kernel density estimation (KDE). KDE is most commonly used to estimate population density and diseases or the incidence of crime (Chainey and Radcliffe, 2005). In this case study the process was used to create a density surface of smoking behaviour, based upon the point locations of postcodes and their attributed health behaviour, to explore the presence of any inherent spatial pattern. It is a useful process for this data because, as noted by Bailey and Gatrell (1995), it produces a smooth estimate of probability density from an observed set of inputs, in this case, the centroids of unit postcodes ascribed with the smoking risks calculated in the previous section.

The KDE process assumes a pattern exists across the study area, not just at the point locations (points associated with code-point polygons, identifying unit postcodes). The technique produces output in the form of a grid, with each cell representing the estimated spatial density of events (Fotheringham et al, 2002; Martin, 1996; Martin **et al.**, 2000). In this case study each cell represents the density of postcodes and their associated attribute values. A detailed description of the process and the formula are documented by Fotheringham and colleagues (2002, pages 146 to 149).

To create the surfaces a kernel was placed over each observation (the spatial coordinates of the postcode centroid) and was then used to spread the attribute values for smoking risk across the study area, according to a given radius. The kernel function was placed over each point in the study region (Greater London) to return a smoothed continuous distribution reflecting the density of postcodes which exhibit certain health behaviours. The resultant surface reflects the intensity of smoking risk at a given point.

To produce the density surfaces, the risk scores of smoking behaviour were transformed into z-scores in order to normalise the distribution (see Harris et al, 2005, page 152) by converting the scores into a standard form. This was done in order to ensure all the survey response variables analysed were brought onto the same scale. It is useful because the index scores are not normally distributed. The minimum index score is constrained by the 0 value, but the maximum value has no constraint, it can continue ad infinitum. An alternative to using z-scores would be to transform the scores using logarithms. The associated z-scores for neighbourhood smoking behaviours were assigned to the points of each code-point polygon (representative of postcode units).

The neighbourhood Types were used as the unique identifier to assign the scores to a spatial reference. A density surface of all smoking behaviours for Greater London was created (Figure 28) using kernel density estimation to highlight patterns and local concentrations of the behaviours. De Smith et al (2007, Chapter 4, online) note 'the choice of grid resolution does not affect the resulting surface, but should be meaningful within the context of the dataset being analysed'. To this end, a grid cell spacing of 75m was chosen because it approximates to the centroid radius of an urban postcode unit, the spatial resolution of the underlying data points. The next step was to choose the bandwidth. The bandwidth controls the amount of smoothing of the density surface, and "larger bandwidths will tend to highlight regional patterns, and smaller bandwidths will emphasise local patterns", (Fotheringham et al, 2002). The purpose of this exercise was to highlight local variations in health behaviour, using a number of tests, the results of which are seen in Figure 27. Here, a bandwidth of 150m was chosen – which appeared to most effectively reflect residential populations and account for holes that represent urban structures such as parks and rivers.

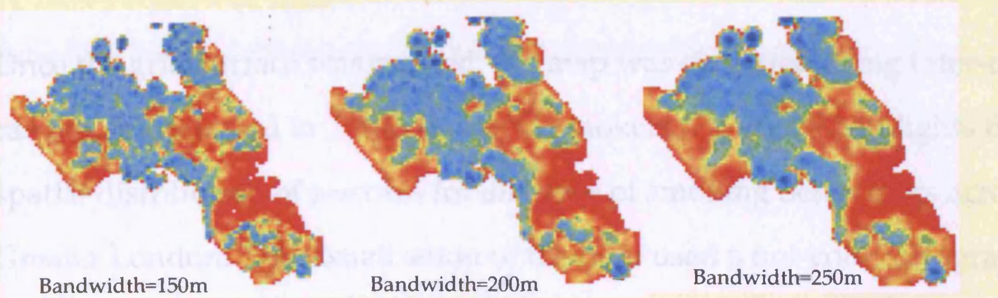


Figure 27: Surface of heavy smokers, made with varying kernel bandwidths

The advantages of using this technique will be discussed in the subsequent case study, but essentially it is used because it facilitates the multiple grids to be combined into composite surfaces of health behaviours.

Using this method it was possible to identify neighbouring geographical locations with high concentrations of the predicted lifestyle behaviour across London. Its purpose was to indicate the relative incidence of neighbourhoods across Greater London most likely to be comprised of cigarette smokers, a useful visual tool for social marketers.

Percentiles	Z-Scores	Index Value	Category
25%	-0.905	Less than 64.5	Likelihood of smoking: considerably below average
50%	-0.182	64.5 to 95.5	Likelihood of smoking: Below average
75%	0.663	95.5 to 131.75	Likelihood of smoking: Above average
100%	2.371	131.75 to 205	Likelihood of smoking: Considerably above average

Table 15: Summarisation of categories used to classify all smokers map

Once the grid surface was created, the map was classified using inter-quartile ranges, summarised in Table 15, for all smokers. Figure 28 highlights the spatial distribution of z-scores for all types of smoking behaviours across Greater London. The visualisation of the map used a hot-cold cartographic technique. Neighbourhoods coloured in dark red represent locations within the top 25% highest smoking z-scores. This corresponded to neighbourhoods where the likelihood of smoking related behaviour is considerably greater than the average for England. Neighbourhoods shaded in light yellow represent locations where the likelihood of residents smoking is considerably below the average for England. Indeed, neighbourhoods coloured in light yellow signify the 25% of London neighbourhoods least likely to smoke.

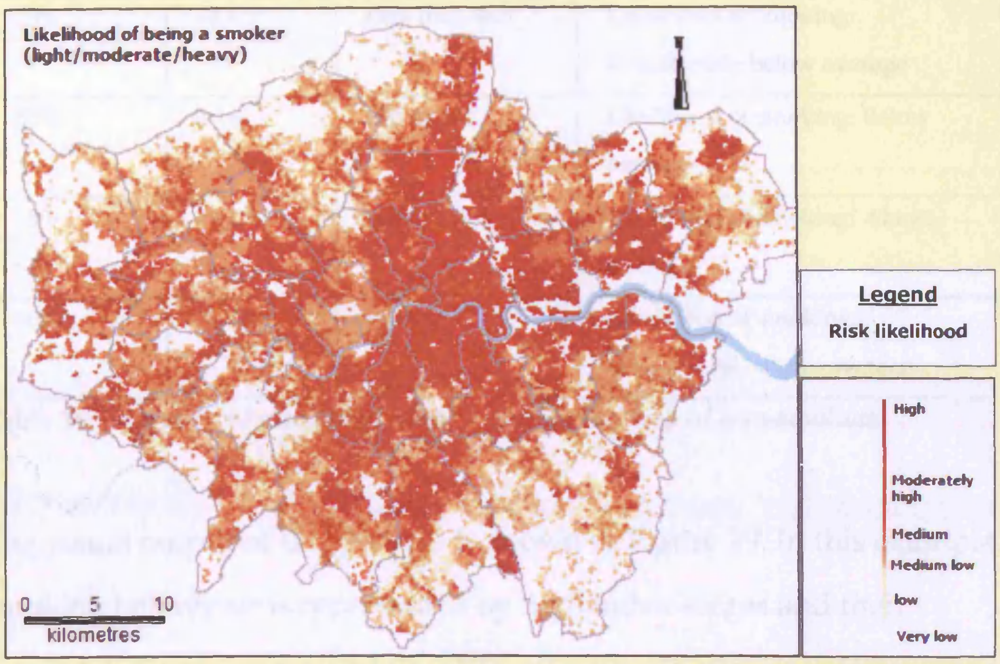


Figure 28: Map of Index Z- Scores for all smokers

The results of the density surface indicated that neighbourhood propensity to smoke appears to be 'clustered' or concentrated. The spatial distribution highlighted neighbourhoods in the London suburbs as being the least likely to smoke and the largest neighbourhood concentrations of smoking related behaviour are predominantly located in east London. Neighbourhoods in north east, east and south east London are most likely to comprise smokers. One limitation is that these surfaces do not provide a measure for the concentration of population.

As should be expected, mapping of the non-smoking variable exhibited the inverse spatial distribution to the all smokers map. Once again z-scores were used to standardise the distribution and mapped using the same technique outlined in the above paragraphs. The output was categorised using quartiles, but the z-score ranges were slightly different, as highlighted in Table 16.

Percentiles	Z-Scores	Index Value	Category
25%	-0.727	Less than 88.5	Likelihood of smoking: considerably below average
50%	0.166	88.5 to 102	Likelihood of smoking: Below average
75%	0.927	102 to 113.5	Likelihood of smoking: Above average
100%	1.622	113.5 to 124	Likelihood of smoking: Considerably above average

Table 16: Summary of categories used to classify the map of non-smokers

The visual output of this process is shown in Figure 29. In this example, non-smoking behaviour is represented by high index scores and their corresponding z-scores. Again a hot-cold visualisation technique was used to categorise the quartile ranges. Neighbourhoods most likely to be non-smokers were categorised in dark red. Neighbourhoods in light yellow, correspond to the 25% of neighbourhoods least likely to be non-smoking. Figure 29 is the inverse of Figure 28, implying spatial concentrations of smoking neighbourhoods across London, as predicted by the Gini coefficients in section 6.3.2. Smoking patterns appear to be spatially and demographically segregated. Further analysis using spatial regression techniques would provide statistical validity to these results, but moves into the realm of explanatory analysis which is beyond the scope of this thesis. The degree to which these neighbourhoods are clustered could be assessed by measuring the level of spatial autocorrelation (section 3.1.1.2), which measures the tendency of similar values to cluster in space. A local Moran's statistic (LISA) would identify and measure the extent of clustering in smoking neighbourhoods.

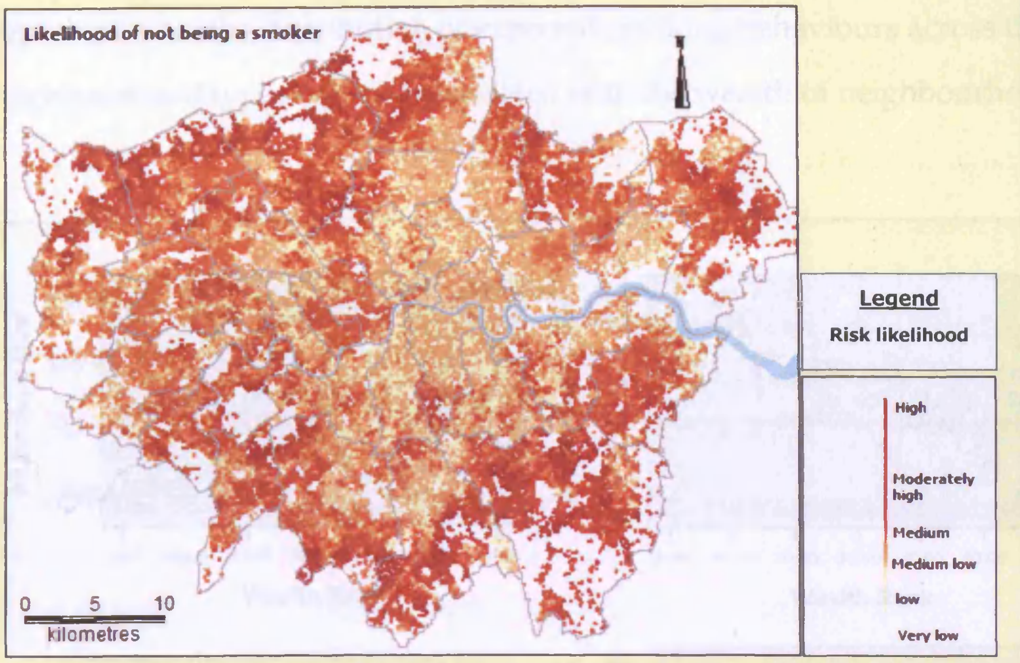


Figure 29: Map of index Z- Scores for non-smokers

6.3.5 Smoking, neighbourhood Types and associations with wealth

In the UK the link between smoking and low income is known (Ash, 2005). So if smoking levels in the neighbourhood can be considered as a suitable predictor of this lifestyle, then they should discriminate between smoking neighbourhoods on the basis of their wealth or income. To reaffirm the validity of exploring smoking behaviours at neighbourhood level the wealth ranks of each Type will be investigated.

Each of the 61 neighbourhood Types used in the geodemographic classification has a corresponding wealth rank. This rank was arranged in descending order; the most affluent neighbourhoods ranked 1 and the least affluent ranked 61. The different standardised smoking indices were plotted against their wealth rank, shown in Figure 30. The association between smoking and income was reaffirmed. The plot of smoking and wealth rank indicated a positive correlation between the two and by corollary lead to the hypothesis that the distribution of expected smoking behaviours across the neighbourhood types was not associated with the wealth of neighbourhoods.

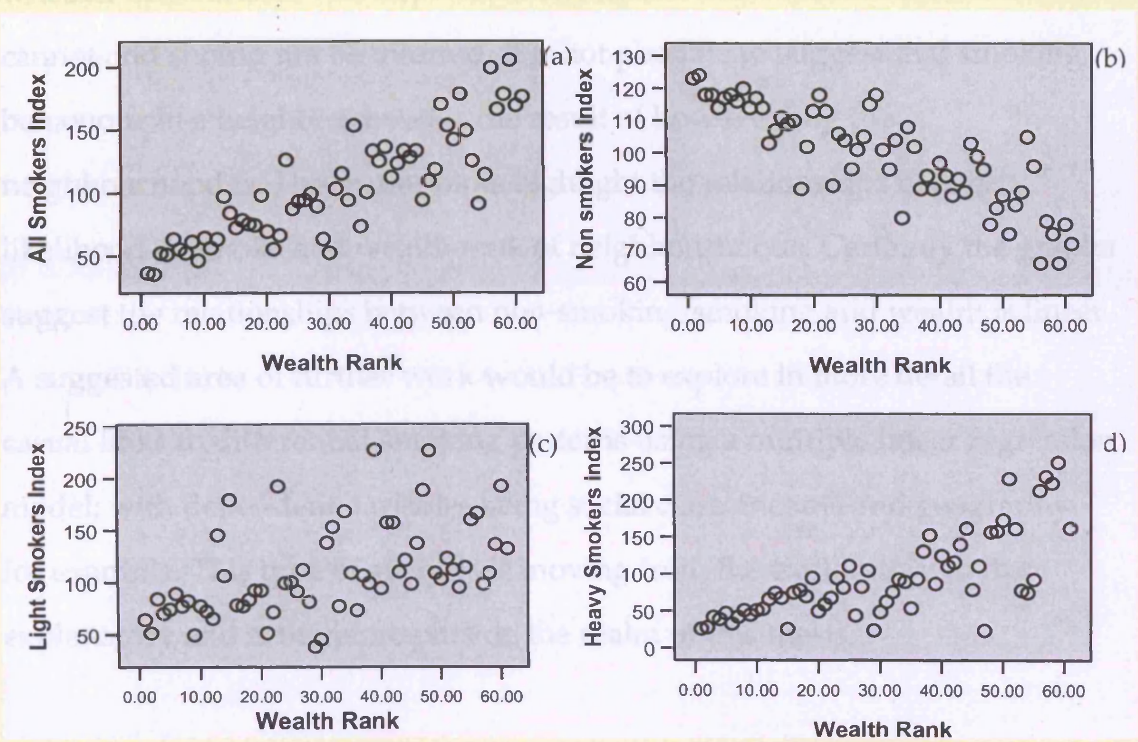


Figure 30: Association of smoking variables and wealth ranks

Once again Spearman's rank correlation was measured for each of the associations in Figure 30. A one-tail test was carried out on these data because the directions of the associations were known. The correlations measured the strength of the predictions made through investigating the standardised index scores; the propensity of neighbourhoods to smoke increases as the wealth rank increases (the rank classifies 1 as the most affluent and 61 as the least affluent therefore as the wealth rank approaches 61 neighbourhoods become less affluent). There was a significant positive correlation for all levels of smokers Figure 30 (a) and their wealth rank, rho = .866, p<.01. There was a significant negative relationship between neighbourhood propensity not to smoke and their wealth rank, rho=-.844, p<.01 (Figure 30, b). Significant positive correlations were also returned for light and heavy smokers and wealth rank but with slightly lower coefficients; rho=.612 and rho=.764, $p<.01$, (Figure 30, c and d) respectively.

Using correlation to assess the relationship makes no priori assumption as to whether one variable was dependent on another and so actual causation cannot and should not be inferred. It is not possible to suggest that smoking behaviour in a neighbourhood is the result of how wealthy the neighbourhood is. The scatter plots highlight the relationships between likelihood to smoke and wealth rank of neighbourhoods. Certainly the graphs suggest the relationships between non-smoking/smoking and wealth is linear. A suggested area of further work would be to explore in more detail the casual links to differential smoking patterns using a multiple linear regression model; with dependent variables being social class, income and geography for example. This type of analysis is moving from the exploratory to the explanatory and is therefore outside the realm of this thesis.

6.3.6 Discussion

Introductory analysis using the Lorenz curve and Gini coefficients identified the variability of smoking behaviours for different neighbourhoods in England, highlighting between group inequalities across neighbourhood Types. Neighbourhoods with greater concentrations of non-smokers are less likely to have high concentrations of heavy or moderate smokers. This observed association in the data was confirmed by using standardised risk index scores to estimate the likelihood of smoking behaviours for different neighbourhoods as compared to the national average. Different smoking behaviours were found clustered within different neighbourhood Types and thus indicate that social marketing strategies for smoking related behaviours should be uniquely tailored and targeted at different populations and smoking behaviours.

Spearman's Rank correlation conducted for non-smoking neighbourhoods and smoking neighbourhoods returned a significant negative correlation. This indicated that non-smokers and heavy smokers do not live in the same neighbourhoods. Smoking behaviours are segregated across neighbourhoods in England and this could be followed up by a spatial regression analysis. Using socio-economic profiles attached to the neighbourhood Types, it was possible to see that smokers were characterised by low income, but also that young transient populations also had a greater propensity to smoke. This association was further explored using the wealth ranks of each neighbourhood. The likelihood of a neighbourhood to be comprised of smokers increased as the wealth rank of the neighbourhood Type increased, i.e. as neighbourhoods become less affluent.

The variation of smoking related behaviour across English neighbourhoods was not the result of random chance, supported by patterns highlighted in the correlation matrix, Figure 26. Smoking behaviour was distributed across a number of similar neighbourhood Types. Local analysis of smoking cessation data provided further evidence of this. The smoking cessation services in the London Borough of Camden are run by Camden Primary Care Trust. Since July 2002 a total of 6971 people have accessed the service. In that time, 3313 people successfully quit, 1569 were unsuccessful and 2079 were lost i.e. stopped attending cessation clinics.

These data were used to identify the types of people who are more successful at quitting. The neighbourhood classification was appended to the client list and a geodemographic profile of successful quitters produced. The results are highlighted in Figure 32. In this instance, the benchmark of 100 was set as the

Camden average; values greater than 100 indicated neighbourhoods more successful at quitting in the allotted four week period.

Neighbourhood Types with values in excess of 100 indicated that clients living in these neighbourhood Types were successful at quitting. By comparison neighbourhood Types below 100 were less successful at quitting and were either lost to the system and require follow-up or did not successfully quit. Almost one quarter of Camden's population live in neighbourhood Type 36. The graph below indicates that their quit rate was below the Camden average, implying that existing services and campaigns were not successful in encouraging this population group to quit. The previous analysis stated that this neighbourhood was 14% more likely to smoke than other neighbourhoods in England. An inequality in need has thus been identified, and is a potential application area for social marketing.

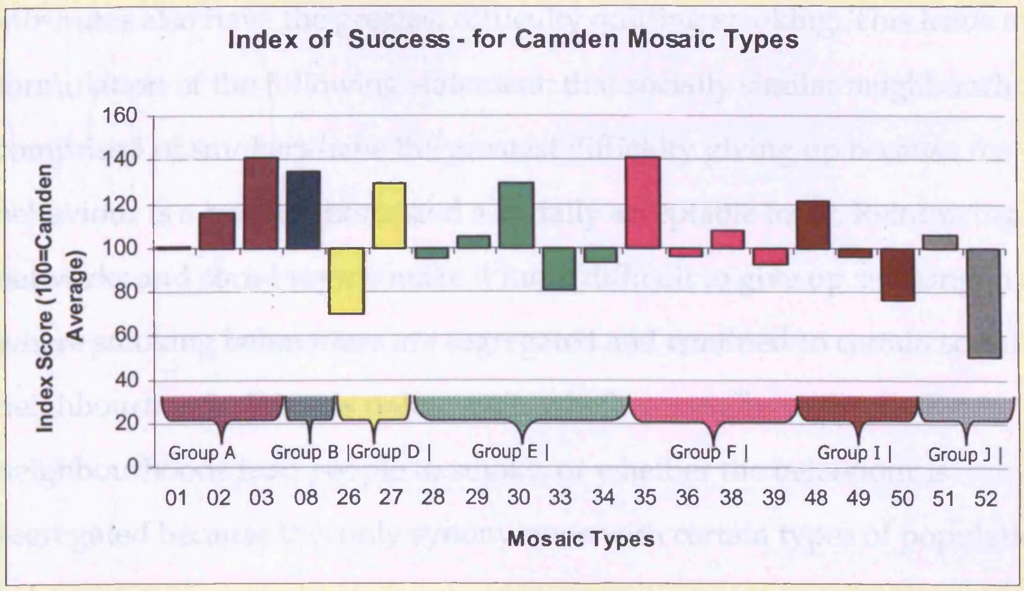


Figure 31: Index of local quitting success for neighbourhoods Types in Camden, London (colours represent hierarchical neighbourhood Groups)

Figure 32 illustrates that Neighbourhoods with greater propensities to smoke have, at least locally in Camden, the most difficulty giving up their addiction. Moreover, the distribution of index scores in Figure 32 does not illustrate any systematic patterning of success in quitting between the members of different neighbourhood Groups. This contradicts what is reasonable to expect, that neighbourhood Types in the same hierarchical group should have broadly similar social characteristics and by corollary have similar chances at successfully quitting smoking. One explanation for this deviation from what is expected, maybe the result of the small size in Camden of some of the Groups. This information combined with the observed patterns in the Health Survey for England data lead to two conclusions. Firstly, a degree of social similarity exists in the way in which smoking behaviours are distributed in the UK. Standardised risk index scores measured observed differences in the dataset and the share of smoking behaviour attributed to neighbourhoods.

Secondly, neighbourhood residents most likely to smoke share similar social attributes also have the greatest difficulty quitting smoking. This leads to the formulation of the following statement: that socially similar neighbourhoods comprised of smokers have the greatest difficulty giving up because the behaviour is a cultural norm and a socially acceptable habit. Reinforcing networks and social norms make it more difficult to give up smoking in areas where smoking behaviours are segregated and confined to certain social neighbourhoods. What is unknown is whether socially segregated neighbourhoods lead people to smoke, or whether the behaviour is segregated because it is only synonymous with certain types of population sub-groups.

Understanding the spectrum of health inequalities across social groups is an important facet of health outcomes research as noted by Marmot and Brunner (2005). In the absence of detailed data for small areas, national surveys of population health can be successfully appended to a geodemographic classification, as shown by this case study. This enables neighbourhood Types to be used as the unit of analysis to conduct detailed exploration into the differential smoking behaviours within and across population sub-groups.

The analysis conducted in this section used 100 as the national average, and calculations were carried out using national proportions of the population to create the baseline data for developing the index scores. The advantage of doing this means the results can be generalised for all postcodes in London and applied to other postcodes outside of London. A more a specific view of the effect of smoking in London, as it is atypical from the rest of the England, could consider using just the population of London to form the denominator in the index score calculation where the average 100 is set to the London average. This would enable neighbourhoods to be compared locally, regionally and nationally, a cautionary not would be that the future statistics maybe limited by the small numbers of some Types which are not represented in London's population. Previous studies of smoking behaviours in the populations support the results and patterns of smoking predicted by geodemographic neighbourhood analysis. Duncan et al., (1999) highlighted the presence of area effects on smoking behaviour which relate to the level of deprivation of the immediate neighbourhood of the individual. The link between smoking and lower social class have been well researched (Peto et al. 2000; Layte and Whelan, 2008, and has been recognised as a major contributor of the 'health-gap' and smoking cessation as a way of considerably reducing inequalities (BBC, 2006).

The use of the neighbourhood is a suitable analysis unit for understanding smoking behaviours because there is a strong link between smoking habits and people's lifestyle and social groups and thus ensures geodemographic classifications provide a useful marker for differentiating. This research shows a new method of exploring local inequalities in smoking behaviour and for this particular behaviour, a geodemographic framework would prove a useful segmentation method for social marketing and a key method in reducing the inequalities that exist between rich and poor in England. Furthermore, if such a technique were used in practice it would enable campaigns to become more efficient, equitable and resourceful by engaging in more specified planning. Those people with the greatest need can be understood with more insight and clarity and specific interventions developed for both the individual and the neighbourhood.

6.4 Predicting alcohol consumption and obesity

Due to the applicability of the neighbourhood for exploring smoking behaviour as explored in the first case study, the second case study builds on this framework to look at the application of neighbourhoods to explore lifestyles related to obesity and alcohol consumption and the how they vary in shared geographic and social space. These two lifestyles were chosen for analysis due to their association with diseases of comfort and their well established relationship with ill health.

6.4.1 Problem definition

In the literature review, Chapter 2, a discussion referenced the many diverse impacts of social determinants on health outcomes. Lifestyle choices of populations, neighbourhoods and individuals all contribute to health-promoting behaviours and outcomes. The previous case study highlighted significant diversity of smoking related behaviours across different neighbourhoods in England. The purpose of this next case study is to expand the thinking on neighbourhood-based lifestyles, to explore the patterns of alcohol consumption and Body Mass Index (BMI) (as a proxy for obesity).

These two lifestyle behaviours were chosen together with smoking because they are on the causal path that leads to a number of chronic illnesses, some of which are diseases of comfort. Being over-weight or obese can bring on lifestyle induced diabetes Type II, a current leading concern for health professionals which is explored in more detail in the case study in Chapter 7. Body Mass Index (BMI) is a ratio of height to weight and is a commonly used measurement to determine if a person is of healthy weight or not. BMI scores of less than 18.5 indicate that a person is under weight, scores between 18.5 and 25 represent healthy weight people and scores over 25 indicate a person is overweight or obese, see Table 10.

Like obesity, excessive alcohol consumption can also lead to long term illness and this is currently the central focus of alcohol policy (such as the national alcohol strategy). Such policy emphasises the social problems caused by excessive drinking and its associations with crime and anti-social behaviour. Very little data are systematically collected on either obesity or alcohol consumption and so this presents a challenge to those wanting to study it. The health survey for England does not measure binge drinking but does monitor alcohol consumption by measuring the number of drinks consumed in the last 7 days, as shown in Table 10. It is impossible to directly infer that people recorded as drinking the most on the HSE are binge drinkers; the results only indicate neighbourhoods that drink more than the national average. It is not possible to make the assumption that one drink is equal to one unit of alcohol, nor is it possible to assume that people are telling the truth when they complete the questionnaire.

6.4.2 Results for weight (obesity) analysis

A number of the variables were aggregated together to create group variables of healthy weight and overweight and obese, as shown in column two of Table 10. It is on these new classified group variables that analysis was undertaken. The outputs derived from the process flow diagram, in 6.2 which created weighted risk index scores that measure the expectation of drinking behaviours and body weight for different neighbourhood Types in England. Figure 32, presents neighbourhood profiles and their associated variability of the risk index scores for those of healthy weight (Figure 32 (a)) and those who are overweight/obese (Figure 32 (b)).

The central value of 100 on the Y-axis is used as the tipping point to represent the national average for England. Values in excess of 100 represent the likelihood of the behaviour being measured to be in excess of the national average. Both charts in Figure 32 show that weight varies by neighbourhood Type, and certain neighbourhoods tend towards being overweight (Figure 32 (b)). The charts are almost the mirror image of each other and unlike the chart for likelihood of failing to stop smoking in Camden (figure 24), there is systematic variation between the groups, as identified by the colour coding in the charts.

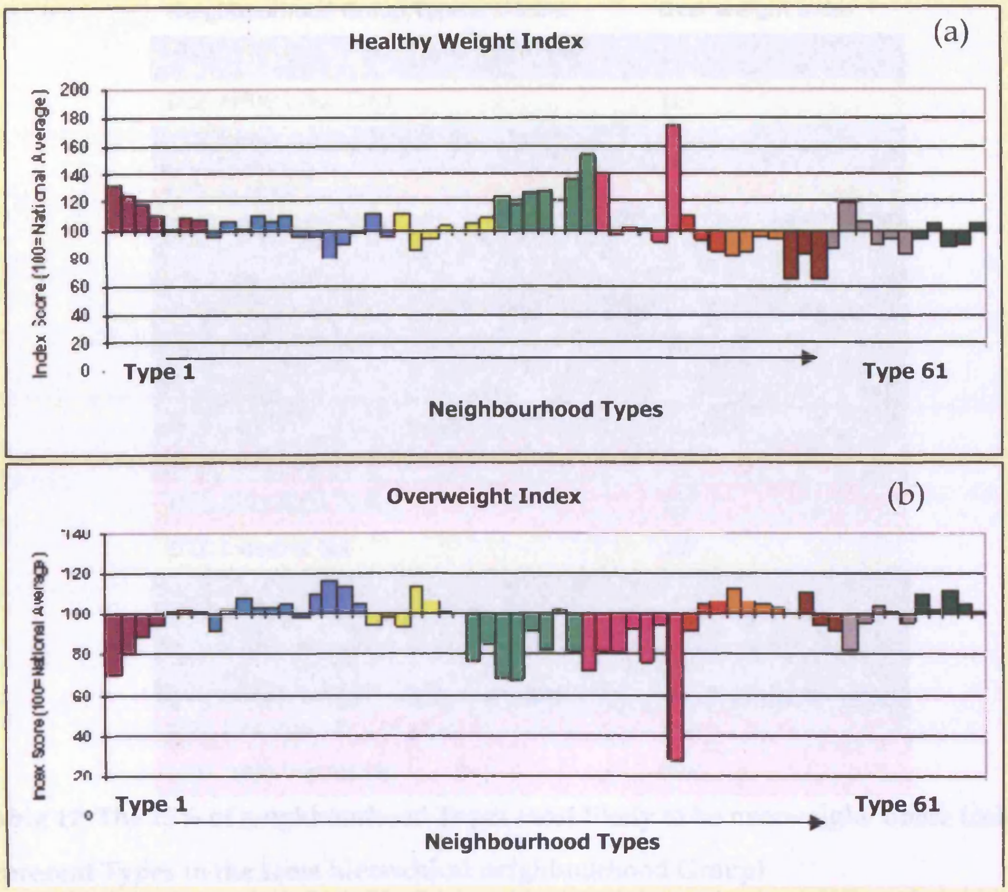
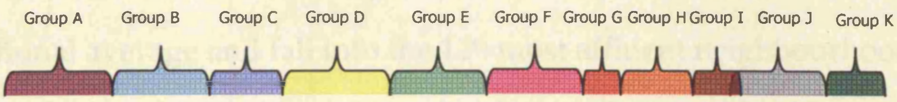


Figure 32: Graphs predicting propensity to be either healthy weight (a) or overweight (b) for neighbourhood Types in England



To identify neighbourhoods most likely to be overweight /obese the data were divided into quartiles. Those with scores greater than 105 were classified as being within the top 25% of neighbourhoods most likely to be obese or overweight, listed in Table 17. An array of neighbourhood Types from a number of different hierarchical neighbourhood Groups are listed in the table. They are coloured according to their hierarchical Group and match the colours used in Figure 32.

Neighbourhood Group,Type and label	Over weight index
C16. Conservative Values	117
D22. Affluent Blue Collar	114
C17. Small Time Business	114
H44. Rustbelt Resilience	113
K59. Parochial Villagers	112
I49. Low Income Elderly	111
K57. Summer Playgrounds	110
C15. Close To Retirement	110
B10. Upscale New Owners	109
G43. Ex-Industrial Legacy	107
H45. Older Right To Buy	107
D23. Industrial Grit	107
K60. Pastoral Symphony	105
B13. Burdened Optimists	105
C18. Sprawling Subtopia	105
G42. Low Horizons	105
H46. White Van Culture	105

Table 17: The 25% of neighbourhood Types most likely to be overweight/ obese (colours represent Types in the same hierarchical neighbourhood Group)

Neighbourhoods representing residents who are more likely to have conservative values are 17% more likely to have BMI scores in excess of the national average and fall into the 12th most affluent neighbourhood Type, according to the wealth rankings in Mosaic. People of healthy weight are most likely to be found in the following neighbourhood Types; Type 33 (Town Gown Transition), Type 34 (University Challenge) and Type 35 (Bedsit Beneficiaries), all neighbourhoods with a young age-profile.

The standard deviations of the index scores are reasonably small. For the overweight/ obese index it is 15.1 which is marginally lower than 19.1, the standard deviation of the healthy weight index scores. These standard deviations indicate the behaviours are more evenly spread across neighbourhoods, than the scores returned by the smoking variables in the previous case study.

The Spearman's rank correlation coefficients for healthy weight versus unhealthy weight returned a significant negative correlation ($p<.001$) with rho=-.670. This can be interpreted as neighbourhoods whose residents have increasingly high index scores for being a healthy weight are not likely to contain residents who are also overweight. Exactly the same technique as that described in section 6.3.4 was used to create the spatial distribution of the variables. The neighbourhood scores were transformed into z-scores. The associated z-scores for neighbourhood weight behaviours were assigned to the centroids of each postcode unit. Once the z-score surfaces were created, the inter quartile ranges of the z-scores were used to classify the maps, outlined in Table 18.

Healthy Weight			
Percentiles	**Z-Scores**	**Index Value**	**Category**
25%	-0.5755	Less than 92.5	Likelihood of healthy weight: considerably below average
50%	-0.1311	101	Likelihood of healthy weight: Below average
75%	0.3394	110	Likelihood of healthy weight: Above average
100%	3.7378	175	Likelihood of healthy weight: Considerably above average
Unhealthy Weight – over weight/obese			
Percentiles	**Z-Scores**	**Index Value**	**Category**
25%	-0.3625	Less than 90	Likelihood of overweight/obese: considerably below average
50%	0.2995	100	Likelihood of overweight/obese: Below average
75%	0.6305	105	Likelihood of overweight/obese: Above average
100%	1.4250	117	Likelihood of overweight/obese: Considerably above average

Table 18: Summarisation of categories used to classify maps of BMI

Figure 33 shows the spatial distribution of z-scores for neighbourhoods most likely to be overweight (a) and of healthy weight (b) across Greater London. In Figure 33 (a) neighbourhoods coloured in dark red represent locations with the top 25% highest z-scores for being overweight/ obese, mapping the spatial locations of the Types identified in Table 18. Neighbourhoods with higher than expected BMI are found in the south, east and west London suburbs and social marketing interventions could be directed at these suburban neighbourhoods. Conversely in Figure 33 (b) neighbourhoods coloured in dark red represent locations with the top 25% highest z-scores for those most likely to have residents with a healthy weight. Neighbourhoods across inner and west London are expected to have a BMI within the healthy weight range (18.5 to 25) and index scores greater than 110.

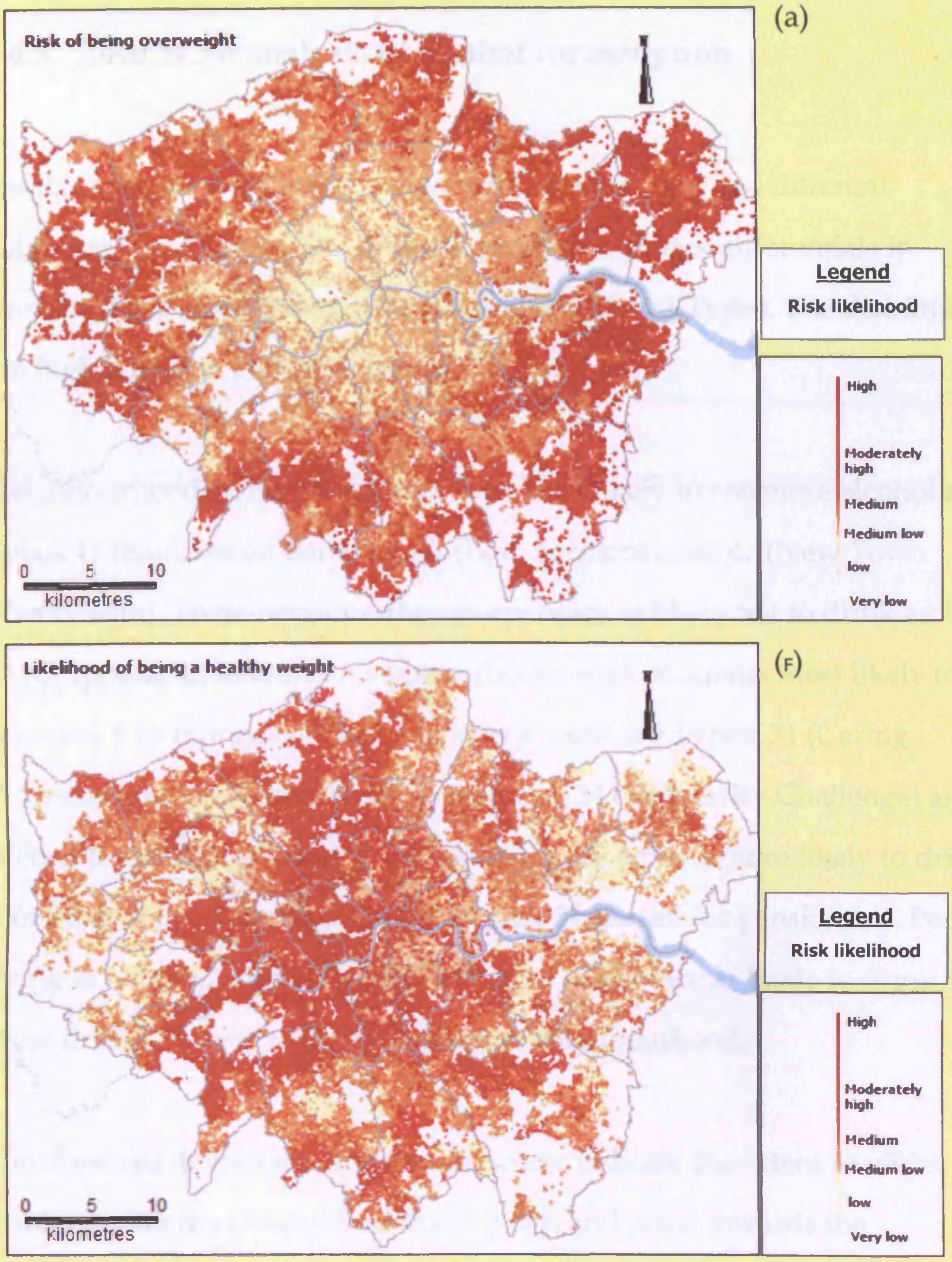


Figure 33: Map of index Z- Scores for neighbourhoods most likely to be (a) characterised by overweight/obese individuals; and (b) individuals of healthy weight

6.4.3 Results for analysis of alcohol consumption

Similar analysis was carried out for drinking habits across different neighbourhood Types. The following section explores differentials in drinking behaviour across different neighbourhood Types. The distributions are highlighted in Figure 34 (a) and (b)

Neighbourhoods where individuals are least likely to consume alcohol are Types 41 (Families on Benefits), 42 (Low horizons) and 47 (New Town Materialism). These neighbourhoods are twice as likely not to drink as Type 03 (Corporate Chieftains). Neighbourhoods with residents most likely to consume 6 or more alcoholic beverages a week are Types; 31 (Caring Professionals), 33 (Town Gown Transition), 34 (University Challenge) and 35 (Bedsit Beneficiaries). Caring Professionals are 6 times more likely to drink more than 6 drinks in one week than type 51 (Cared for pensioners). People living in neighbourhood Type 31 are more than twice as likely to drink 6 or more drinks in a week, than the average neighbourhood.

The standard deviations of the index scores indicate the extent to which drinking patterns of neighbourhoods vary, and point towards the discriminatory power of the neighbourhood to differentiate lifestyle choices. Neighbourhoods where residents are most likely not to drink have a standard deviation of 26.1, 1 to 3 drinks = 27.9, 3 to 6 drinks 26.6 and the standard deviation for neighbourhood scores for more than 6 drinks = 32.5. The scores are marginally higher than those seen for the BMI weight variables but lower than the standard deviations for smoking variables. The drinking variable with the most discriminatory power is that which identifies individuals who consume more than 6 alcoholic beverages within seven days.

Spearman's rank correlation coefficients matrix for drinking related behaviours	No Drink Index	Between 1 and 3 drinks Index	Between 3 and 6 drinks Index	More than 6 drinks Index
'No Drink' Index		-.442	-.634	-.039
Between 1 and 3 drinks Index	-.442		.611	-.471
Between 3 and 6 drinks Index	-.634	.611		-.135
More than 6 drinks Index	-.039	-.471	-.135	

Table 19: Spearman's rank correlation coefficients matrix for drinking related behaviours

The Spearman's rank correlation coefficients for the various drinking related behaviours are summarised in Table 19. A number of these correlations are statistically significant at p=.001, meaning there is less than a 1 in 100 chance that the distributions are the result of random chance. Neighbourhoods where residents are likely to drink more than 6 drinks a week are less likely to have residents only drinking between 1 and 3 drinks. A similar negative association exists between neighbourhoods where residents do not drink and neighbourhoods where residents drink 3 to 6 drinks in a week. These results are statistically significant patterns and highlight a degree of association for different drinking behaviours.

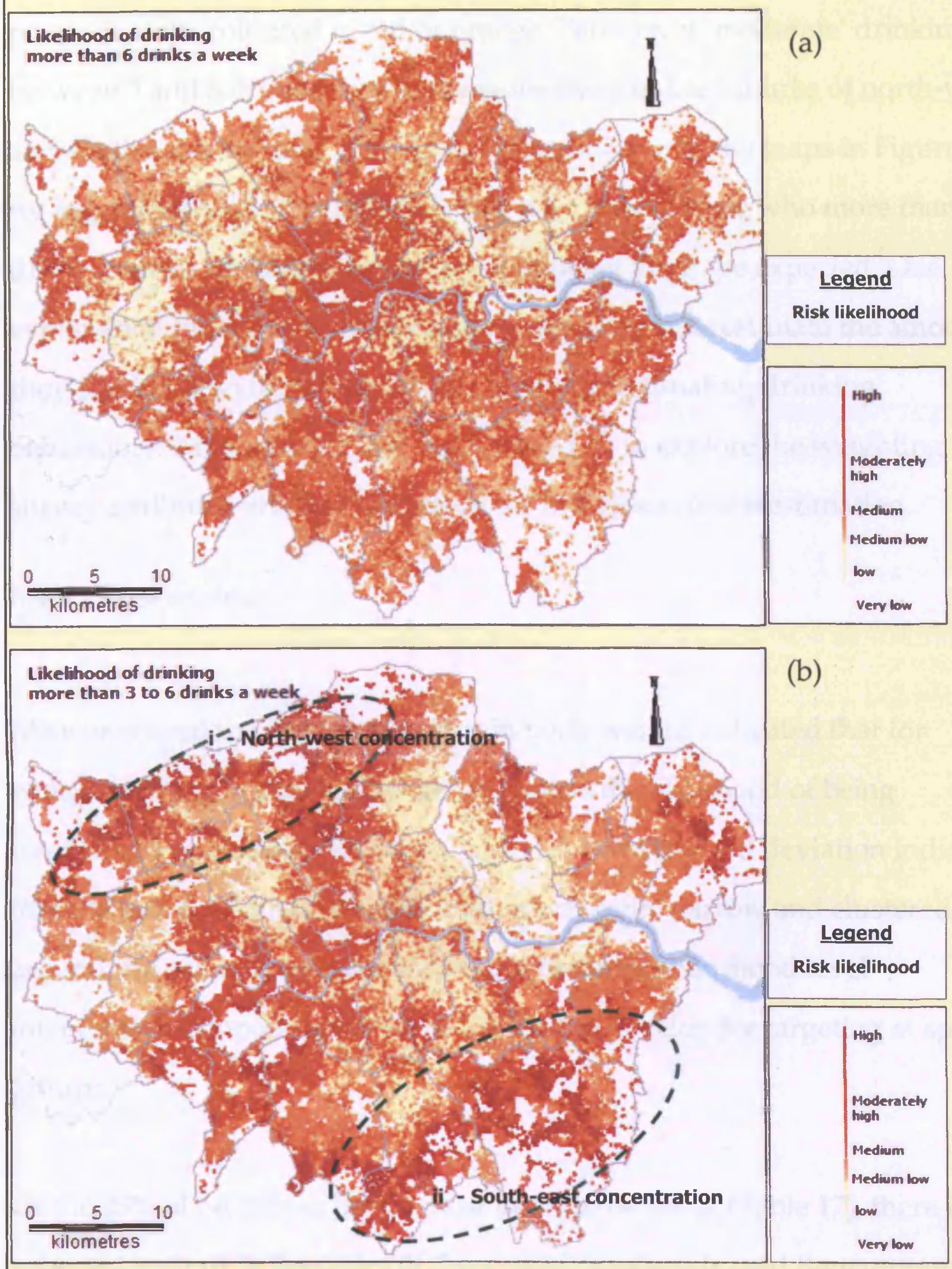


Figure 34: Neighbourhoods (a) where residents are most likely to consume more than 6 alcoholic beverages and (b) where residents are most likely to consume more than 3 and 6 beverages per week

Mapping of the drinking variables can be used to illustrate the spatial variability in alcohol consumption across London, as shown in Figure 34 (a) and (b). Neighbourhoods where individuals consume more than six drinks a week are predominantly located in the west London and the suburbs where the maps are more orange, and can be identified in Figure 34 (a) by the postcode units coloured in red or orange. Patterns of 'moderate' drinking, between 3 and 6 drinks a week, are more likely in the suburbs of north-west and south –east London. The visual differentiation of the maps in Figure 34, for neighbourhoods most likely to drink 3 to 6 and those who more than 6 drinks a week appears to be less red than what might be expected. One explanation for this is the tendency of people to underestimate the amount they drink and so the survey counts are underestimating drinking behaviours. An area of further work would be to explore the weighting of survey attribute variables to account for individual underestimation.

6.4.4 Discussion

Measures used to predict variability in body weight indicated that for residents living in different neighbourhoods the likelihood of being overweight or obese varies only slightly. A low standard deviation indicated that the actual width of the distribution was quite narrow and clustered around the national average, indicating that neighbourhood level intervention campaigns would be difficult to develop for targeting at specific groups.

Of the 25% of neighbourhoods most likely to be obese (Table 17), there is no coherent pattern to the order of these neighbourhoods, and they represent some of the most affluent and the least affluent neighbourhoods, indicating that unlike smoking, obesity is not simply a disease associated with poverty and low income. This suggests that obesity is not a disease of the rich or the poor, as further supported by the scatter plot in Figure 35 and there are significant results for Spearman's rank correlation analysis of wealth rank and the overweight/obesity index.

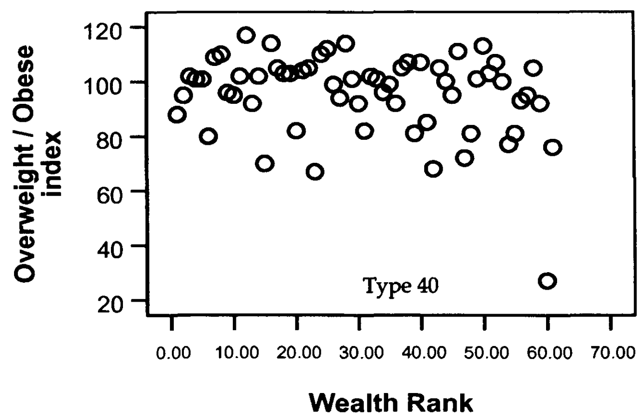


Figure 35: Association of weight variables and wealth ranks

The point marked as Type 40 in the graph appears to have a very different score compared to the other neighbourhoods. It is an outlier in the dataset and has a very high wealth rank – indicating a low income Type but also is the neighbourhood most likely not to have obese or overweight residents. It represents a neighbourhood mostly found in Scotland, and therefore identifies a limitation of using a UK wide geodemographic classification to assess the results of an England survey.

Other geographical studies about obesity have been conducted by various researchers across the UK. A recent article in 2005 explored the distribution of fast food outlets in deprived neighbourhoods, the thesis of the paper being that the presence of such food outlets contributes to the greater prevalence of obesity in deprived neighbourhoods (Cummins et al., 2005). The paper linked the number of McDonalds restaurants to IMD super output areas to conclude that there are more of them located in deprived neighbourhoods; the results do little other than indicate certain locations have more restaurants than other areas and the results are subject to the ecological fallacy. Furthermore, this research based its notion of obesity and deprivation on old studies by

Ellaway et al. in 1997 which observed 600 people living on the west coast of Scotland, it is probably another ecological fallacy to take the results of a study in Scotland and apply in more generally, which is why the results of this Scottish study differs to that of this thesis. In 2007 Moon et al., developed a complex picture of obesity using multi-level synthetic estimates for different age, sex and ethnicities derived from the Health Survey for England. The study concludes that the geography of obesity is not the same for everyone everywhere (page 29), but is more complex than has been assumed from previous evidence. The complexity of obesity in the population was highlighted in the research in this thesis by the lack of a clear differentiation between different neighbourhood Types and Groups and considers the variation in local distributions are not simply linked to socio-economic factors.

The density maps created just for London indicated a tendency of neighbourhoods with overweight residents to be represented by postcodes located in the suburbs of London. This may signify a relationship in the data not adequately explained or explored by geodemographics alone, perhaps linked to car ownership, or could be the result of the neighbourhood Types derived from a national classification that happen to be found in London. What this does indicate is that for some variables micro-level analysis for small neighbourhoods is not sufficient at explaining or exploring variant patterns. In an analysis of ethnicity and segregation, Johnston and colleagues at Bristol University (2006) found that the functional scale of the school was better at identifying segregation patterns than the functional scale of the neighbourhood. It might be that other functional units besides neighbourhoods are better discriminators of obesity, an area of research worthy of further work. So whilst it is beyond the scope of this study to move into explanatory analysis, further research would be required to explore the reasons for these patterns and their significance; such research would include local indicators of spatial association. The extensions of this research are discussed in more detail in the further work chapter, but point toward carrying out a geographically weighted regression to explore variability in health related outcomes at inter-urban scales of analysis.

This case study has built upon the framework presented in the previous example to explore the way in which geodemographic techniques can contribute to the debate around differential health outcomes across different populations, demonstrating that aggregation of data according to social similarity is a useful tool for understanding and describing patterns of health behaviour. Neighbourhood patterns of drinking and obesity are less segregated than smoking behaviours, suggesting that the inequalities seen in the distribution of these outcomes are not simply related to socio-economic contexts.

6.5 Composite spatial indicators of lifestyle

6.5.1 Problem definition

For improving health initiatives and campaigns it is pertinent to investigate the associations between each of the three behaviours explored in the previous analysis, which could lead to a more joined up approach to resolving health-harming behaviours. A common technique in policy making is to merge many different data variables to create one composite indicator of need. This next section explores a geographically centric approach to joining the data, in order to create a composite map of health behaviour.

6.5.2 Method, analysis and results

Exploratory analysis was undertaken to consider the associations between the most healthy, moderately healthy and unhealthy lifestyle behaviours. Healthy neighbourhoods were defined as neighbourhoods where residents tended not to smoke, consumed less than 3 drinks (although 1 drink cannot be equated to 1 unit of alcohol) a week and were of healthy weight. Moderately healthy neighbourhoods were classified as having residents who tended not to smoke, consume between 3 and 6 units of alcohol a week and have healthy a weight. Unhealthy neighbourhoods were comprised of residents likely to smoke, consume more than 6 units of alcohol a week or be overweight or obese.

Transformed index scores were used to map the newly created composite lifestyle variables; healthy, moderately healthy and unhealthy neighbourhoods. The transformed scores allowed comparisons across the range of variables, indicating how different a neighbourhood is from the average. The transformed scores enabled the aggregation of the three indices to produce a graduated scale of neighbourhoods exhibiting unhealthy, healthy and moderately healthy behaviours.

The density surfaces of z-scores derived from the previous case studies were used to generate new information. The grid calculator in MapInfo's Vertical Mapper was used to apply spatial algebra to the grids, creating new spatial surfaces of the three composite health behaviours; healthy, moderately healthy and unhealthy neighbourhoods, see figure below.

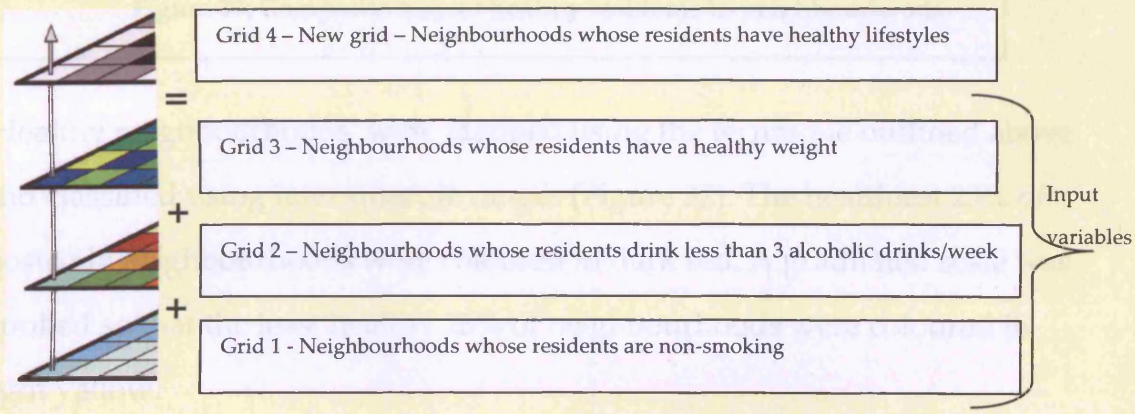


Figure 36: Grid algebra used to calculated composite indicators of lifestyle characteristics for residents in a neighbourhood

Figure 36 breaks down the method used to calculate the derivative map of healthy neighbourhoods. It is the sum of the transformed scores for the 3 input variables. The new surface (Figure 37) created is representative of a composite healthy lifestyle variable. Neighbourhoods with large scores were indicative of locations where residents were expected to have the mostly healthy lifestyles.

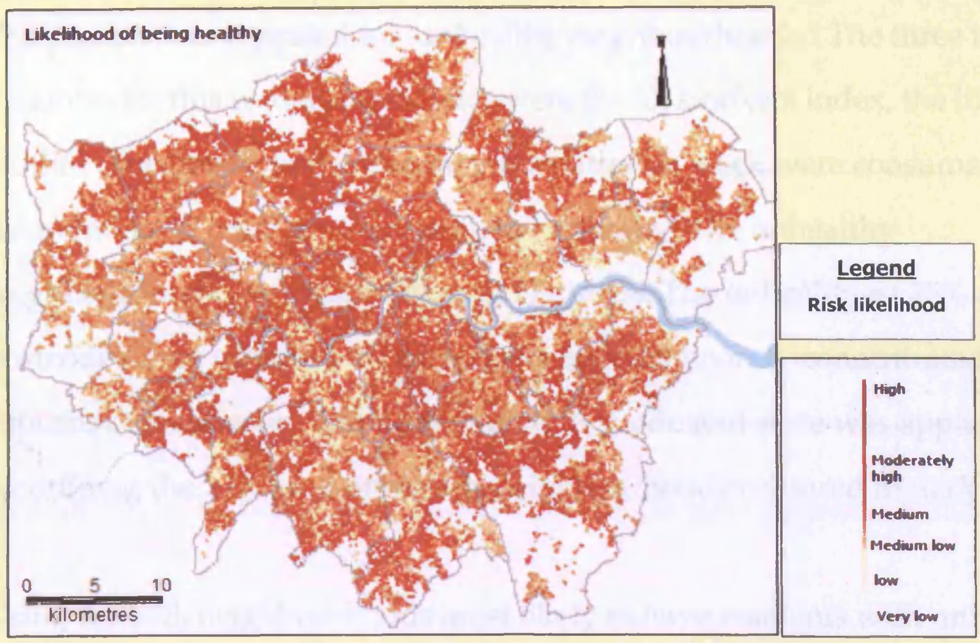


Figure 37: Composite map of healthy residents in neighbourhoods

'Healthy neighbourhoods' were mapped using the technique outlined above and classified using inter-quartile ranges (Figure 37). The healthiest 25% of postcode neighbourhoods were coloured in dark red. A graduated scale was applied so that the least healthy 25% of neighbourhoods were coloured in light yellow.

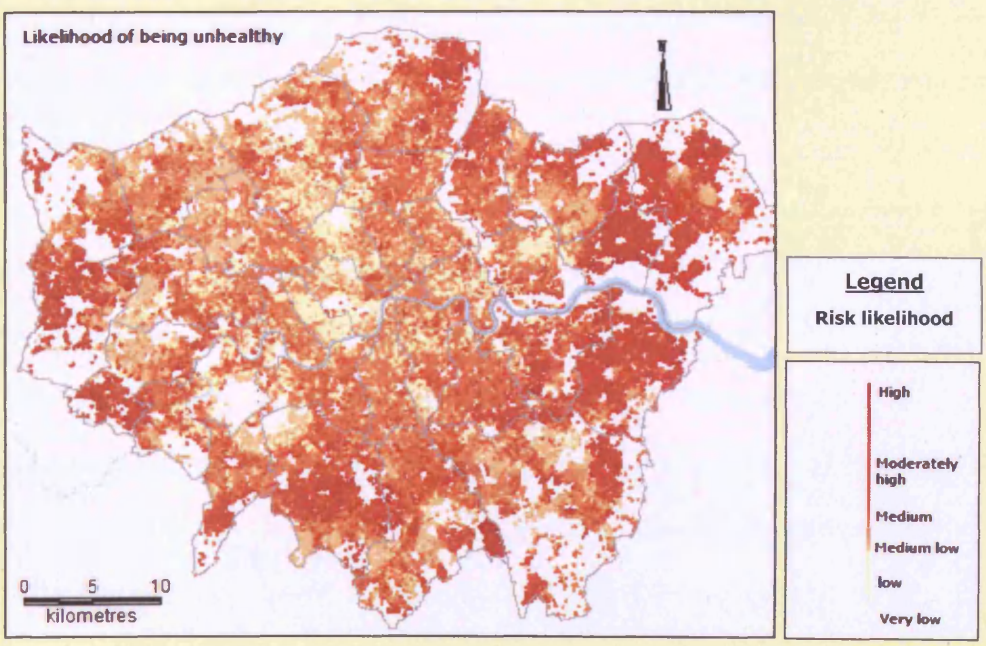


Figure 38: Composite map of neighbourhoods whose residents are unhealthy

The process was repeated for 'unhealthy neighbourhoods'. The three input variables for this composite indictor were the all smokers index, the index for alcohol consumption where more than 6 drinks a week were consumed and the index score predicting overweight or obesity. The unhealthy neighbourhoods were mapped, as in Figure 38. The unhealthiest 25% of postcodes were coloured in red and were predominantly concentrated in the suburbs of west, south and east London. A graduated scale was applied identifying the 25% least unhealthy neighbourhoods coloured in dark blue.

Using the GIS, neighbourhoods most likely to have residents with unhealthy behaviours were extracted to create a new grid, representing the 25% most "unhealthy" neighbourhoods in London, see Figure 39.

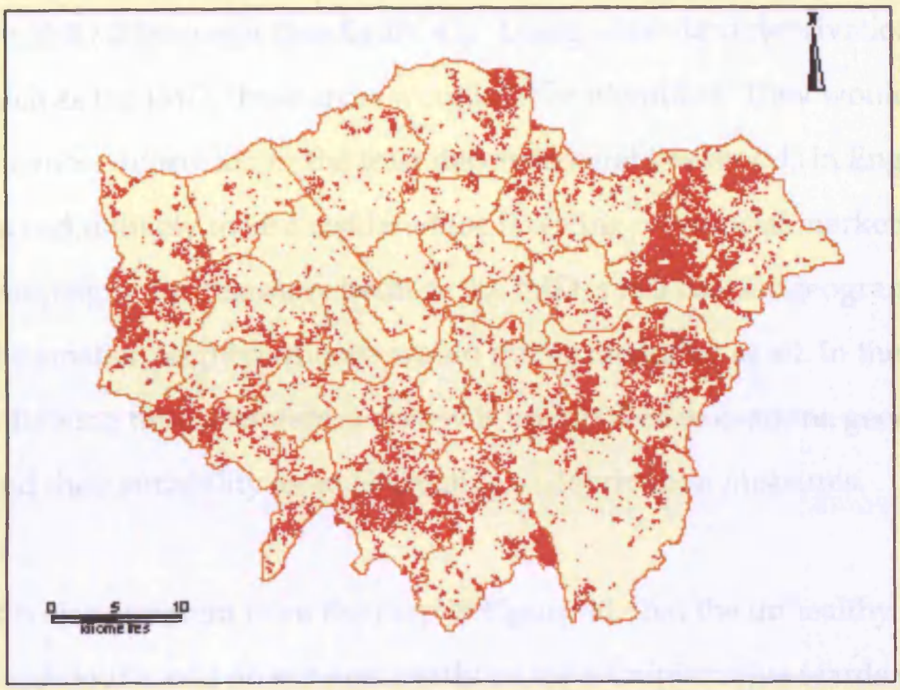


Figure 39: Locations of the 25% most unhealthy neighbourhoods in London, predicted using a composite indictor of health-harming lifestyle behaviours

For the purposes of social marketing, having a local understanding of behaviours and conditions is essential and fits with the understanding of place that comprises one of the four ps of marketing. Figure 40 zooms in on Camden, to highlight the neighbourhoods that fall within the 25% most unhealthy in London. Residents in these neighbourhoods are more likely to smoke, be overweight and consume more than 6 drinks a week compared to other neighbourhoods in Camden.

The main map, Figure 40, highlights the neighbourhoods in Camden that fall into the 25% unhealthiest neighbourhoods in London, and are coloured in red on the map. Neighbourhood renewal areas, localities set aside for regeneration, are marked in an orange cross hatch. Some of the unhealthiest neighbourhoods in Camden do not fall into these regeneration areas. Furthermore the 'cluster' /'hotspot' found in the southern borough of Bloomsbury does not rank highly using the Index of Multiple Deprivation

(IMD). The comprising super output areas fall in the inter-quartile range of the IMD 2004 scores (see figure 41). Using a standard deprivation measure such as the IMD, these areas would not be identified. They would neither be the most deprived nor the least deprived neighbourhoods in England so would unlikely to be considered for targeting with social marketing campaigns. Furthermore because the IMD is at a coarser geographical scale the smaller neighbourhoods would not be identified at all. In the discussion following these chapters, a review is undertaken to examine geodemographic and their suitability as an alternative to deprivation measures.

It is also apparent from the map in Figure 40, that the unhealthy neighbourhoods do not nest neatly inside administrative wards (marked by the blue line) and so these boundaries should not be used to plan interventions within. This highlights the issues related to scale and the impact of aggregations to arbitrary administrative boundaries, an issue returned to in Chapter 8.

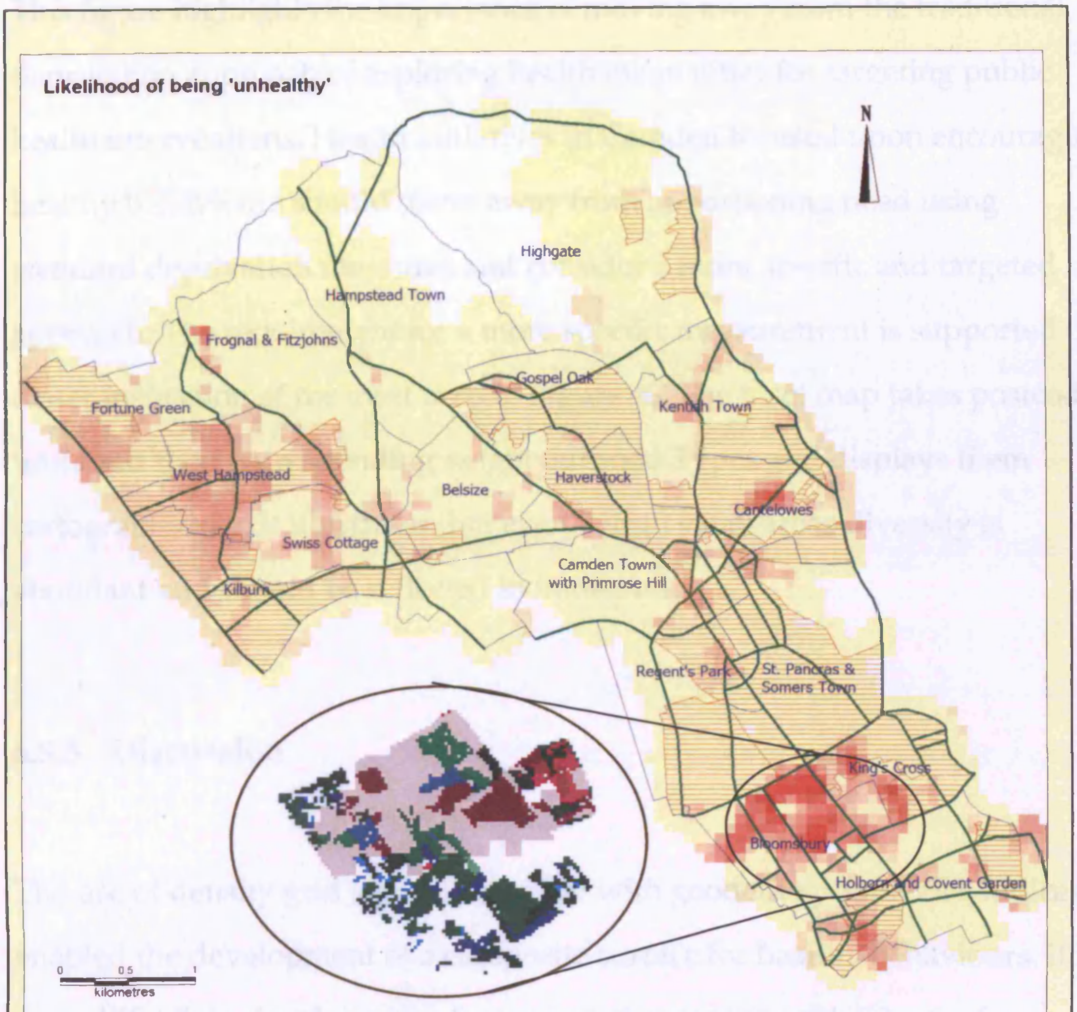


Figure 40: Neighbourhoods with the unhealthiest residents, in Camden, London.

This figure highlights the importance of moving away from the traditional deprivation approach of exploring health inequalities for targeting public health interventions. Health initiatives in Camden focused upon encouraging healthy behaviours should move away from apportioning need using standard deprivation measures and consider a more specific and targeted approach. The requirement for a more specific measurement is supported on closer inspection of the inset map in Figure 40. The inset map takes postcode units and their corresponding neighbourhood Types and displays them cartographically. It illustrates that even within small areas diversity is abundant and should be reflected in initiatives.

6.5.3 Discussion

The use of density grid analysis together with geodemographic risk indices enabled the development of a composite surface for healthy behaviours. It is very difficult to develop a single measure that captures all aspects of a complicated and large dataset. The technique used in the above case study is most suited to continuous data but despite this, the results illustrated the usefulness of creating a composite geographical variable of health behaviours to explore disparities across London. Further work should consider other techniques; such as applying weights to the different grid surfaces to give one lifestyle a stronger emphasis (such as smoking) in the composite indicator. The composite surface as developed in this chapter assigns equal weights to the grids during the map algebra. It was decided to assign equal weights as it was difficult to determine which should be given more weight. It may even be that for different areas different weights should be applied according to the population characteristics of the local area.

6.6 Chapter summary & discussion

The complex nature of health inequalities and the identification of local health needs present many challenges to primary health care: social marketing, commissioning, planning and practice. Routinely, health needs are measured according to administrative units within which people live, but this measurement system in itself is limited because the units of analysis merely act as containers for the phenomena being measured (Harris and Longley, 2002). Consequently, the vision of policy at the local level is often limited to focusing on the most deprived neighbourhoods, as measured by the Index of Multiple Deprivation (Figure 41) or neighbourhoods that have been identified for some other area-based policy, such as neighbourhood renewal. This limits thinking to specific domains of health care planning to coarse zones (such as Super Output Areas), as they are not available at finer spatial areas, or area schemes that have little to do with health care provision or the reason for the intervention. If the patterns of deprivation highlighted in Figure 41 are compared visually (bearing in mind they are produced at different scales) to patterns of heavy smoking, high alcohol consumption and drinking and obesity, the only attribute that follows a similar spatial distribution is that of heavy smoking (Figure 24), which suggests that the distributions are not always associated to deprivation and are indeed diseases of comfort.

A fundamental problem with most health care planning is that datasets (such as the HSE) which could present a rich and detailed picture of health related behaviours are collected using sample designs that do not permit inference at the fine levels of spatial granularity that would be most helpful to health care professionals. Also there is a heavy reliance on the IMD 2004 (see Figure 41) because of the association between poor health outcomes and deprivation. The only data sources that are available at such spatial scales record only a restricted range of attributes, that may be appropriate for identifying poor physical and social conditions at, say, the scale of the Super Output Area, but do not indicate the full spectrum of specifically health related behaviours at these and finer scales.

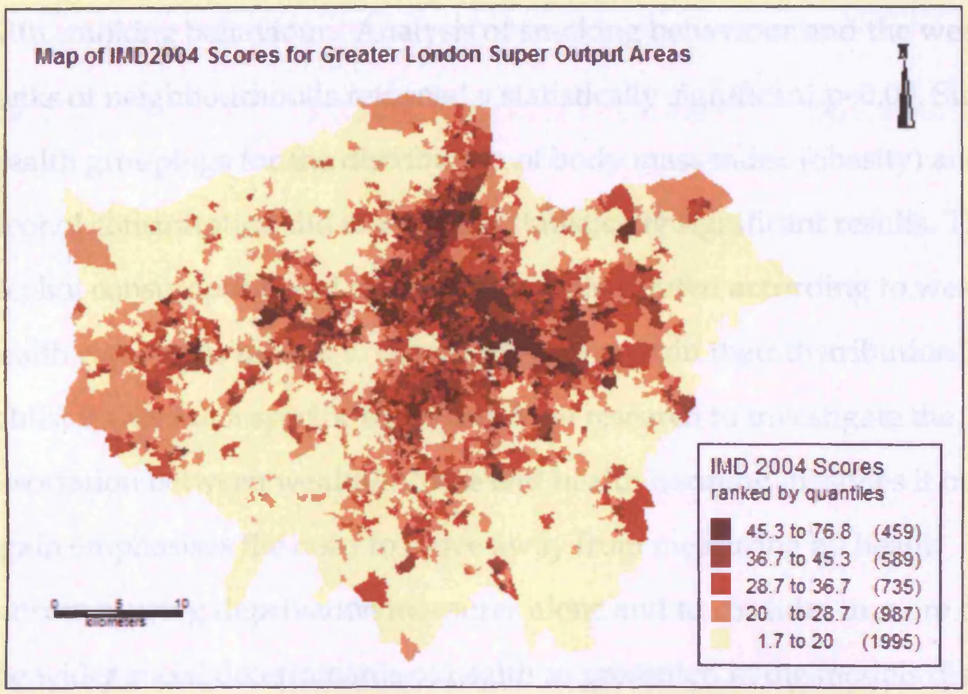


Figure 41: IMD 2004 for London at the level of Super Output Areas

It is in this context that these case studies have focused upon the task of inferring the local geography of a broad range of health care needs and priorities, using a geodemographic classification to link the characteristics and attitudes of survey respondents to the attributes of small neighbourhood areas. The results of the research suggest differences in the ways in which the three health behaviours vary between neighbourhood demographic Types. The use of geodemographic classifications stratified the data across different socio-economic groups that accounted for local and regional population differences. Geographic visualisation of the results and further statistical analysis illustrated the patterns in the data were not simply the result of socio-economic differences

The density maps identified local concentrations of likely health impacting behaviour on residents. In the case of smoking, discerning patterns of this specific health behaviour were highlighted using geodemographic Types to segment the dataset which also reinforced the income inequalities associated with smoking behaviour. Analysis of smoking behaviour and the wealth ranks of neighbourhoods returned a statistically significant $p=0.05$. Similar wealth groupings for the distribution of body mass index (obesity) and alcohol consumption did not display statistically significant results. Thus, alcohol consumption and obesity are not distributed according to wealth, and health care needs to look to other factors to explain their distribution. So whilst it was not a specific objective of the research to investigate the association between wealth/income and health harming lifestyles it once again emphasises the need to move away from measuring all health outcomes using deprivation measures alone and to consider in more detail the wider social determinants of health as presented in the models discussed in Chapter 2. The replication in the research of the established link between wealth and smoking activity at the neighbourhood level provides some validation of the method.

Using this alternative approach to exploring patterns of health needs makes it possible to incorporate a range of socio-economic and lifestyle considerations into the analysis, encompassing the many different social determinants of health into one classification. The geodemographic classifications already account for socio-economic distributions and thus using them to segment and classify survey removes the need for many auxiliary social data variables. Combining the classification with health survey data to produce neighbourhood quotients, local and regional comparisons were attainable.

The implications may be significant for local policy implementation, resource allocation, service distribution and social marketing. If conventional policy implementation is centred on neighbourhoods that are traditionally described as deprived, then for certain health behaviours evident across social and wealth divides (such as obesity), interventions will not be sensitive to health behaviour distributions and will not reach those in need. There is a requirement to move beyond the bottom 10% approach of policy implementation, which selects the 10% (sometime 25%) most deprived (using IMD) neighbourhoods to be considered for interventions. This approach clearly does not always identify population sub-groups in need as highlighted in the smoking example. Expansion of this thought will be continued in the discussion section of the thesis (Chapter 8), which discuses the appropriateness of geodemographics as a viable alternative to deprivation methods.

The health-related behaviours displayed by individuals cannot be understood without taking account of the characteristics of and the processes occurring at the levels of both the immediate and broader environments (Macintyre and Ellaway, 2000, page 337). As identified by Berkman and Kawachi (2000, page 7) most behaviours are not distributed randomly and they are socially patterned and often cluster with one another. Whilst smoking, drinking and weight gain are individual health risk factors the aggregation of these variables using a geodemographic classification enabled the development of neighbourhood level predictors.

Geodemographics encapsulate lifestyle factors relating to smoking and are a useful neighbourhood discriminator for this lifestyle behaviour. In contrast, they are less able to differentiate local patterns of diet and alcohol consumption, which are more complex to identify. It is possible to speculate that these differences in indictor efficiency arise because smoking behaviours may be linked to social norms and social capital. Thus for some sub-groups of the population smoking is an acceptable behaviour, and it those neighbourhoods with high levels of homogenous bonding capital (section 2.4.3) that help to reinforce the health harming behaviour amongst the social group. Further exploration into these differences could be achieved using a qualitative approach, perhaps through in-depth interviews, to enable a more concrete evidence base.

Chapter 7

Geodemographic and health outcome analysis

Geodemographic and preventative health outcome measurements informing existing public health services

7 Geodemographic and preventative health outcome measurements informing existing public health services

The previous chapter illustrated how geodemographic techniques can be applied to national survey data relating to health to explore, at the local scale, differential patterns of lifestyle behaviours with health- harming influences. The results of which indicated that of the three behaviours explored smoking behaviour was best discriminated at the neighbourhood level. This chapter substantiates the methodology and considers how the techniques can be developed in a real world public health setting, to provide the robust indictors necessary to create an evidence base for local decision-making around preventative care and social marketing. The purpose of this chapter is to explore the development and realisation of the second objective outlined in the research methodology (Chapter 4. section 4.5)

Objective Two

Substantiate health variability indicators using framework in real world setting, to enable investigation of scalability and validity of measures

Once again the absence of good quality, accurate local data inhibits the development a comprehensive picture related to public health issues such as diabetes and screening uptake, together with other potential diseases of comfort. The consequence of the unavailability of good quality data means it is more difficult for public health professionals to consider innovative and imaginative initiatives for encouraging behaviour change. Of the three case studies presented in this chapter, two investigate the capability of geodemographic analysis to be used for informing diabetes and breast screening initiatives while the other study considers how the data can be usefully ascribed to general practitioners' patient lists to develop practice profiles. Each of the case studies presents the indicators at a variety of different spatial scales, which illustrates their versatility to a wide range of scenarios and illustrates their scalable nature, particularly mindful of the decision-making structure of the NHS. The first case study in this chapter also attempts to validate the technique by comparing actual versus predicted diabetes outcomes in order to determine the extent to which the accuracy and robustness of the indicators can be assessed.

The conditions being explored in this chapter were chosen because of their strategic importance within the health care setting. Diabetes falls into the category of diseases of comfort because its causal risk factors are associated with health-harming behaviours such as obesity and poor diet. For this reason, the first case study in the chapter looks at how existing transactional operational data connected to diabetes related hospital admissions can be combined with geodemographic classifications to develop indicators which predict the likelihood of being admitted to hospital for diabetes related conditions. These indicators are then applied to a range of geographical and social scales to provide health practitioners with information which will assist their planning of interventions, campaigns and possibly even services.

The measures are assessed in relation to their scalability to formal administrative areas, by creating aggregate measures for the boundaries designating the responsibility of Primary Care Trusts (PCTs). Furthermore the results of the analysis are compared with a nationally accepted model for predicting diabetes to both substantiate and ascertain the validity of the method used in this thesis.

The second case study in this chapter takes the principles used to create the indictors of diabetes needs and extends them to aid the creation of practice profiles for Camden based general practices and to predict two types of diabetes conditions. The neighbourhood is most associated to the social scale of the residential population which is considered as the functional scale (Moon, 1990), but decision making and targeting in the NHS is often conducted at larger scales of analysis (see chapter 1, page 7, Longley et al., 2001). Thus, it is important to consider the more formal scales of the NHS decision making process, if the framework is to be successfully implemented. Additionally, social marketing campaigns are not simply restricted to neighbourhood targeting and PCTs have the potential to target people – in conjunction with general practices, for example. Therefore, the flexibility, adaptability and scalability of indicators reviewed in the first case study of this chapter is important. These notions are extended to encompass their application to functional scales (the geography at which care is delivered) through a study exploring the practice profiling of risk likelihood. This case study is conducted for a range of diseases and conditions to determine if the developed framework is flexible and robust enough to encompass such a wide range of diseases/conditions.

The final case study in this chapter draws together all the research outcomes in the first two case studies of this chapter and the previous chapter. The concluding case study in this chapter consolidates all of the research knowledge acquired and was then applied within Camden PCT to explore the application of the neighbourhoods as a predictor for response to breast screening services. In the final case study operational health data are used to measure and predict service uptake.

7.1 Predicting diabetes health risk

7.1.1 Problem definition

Diabetes is a chronic long term illness responsible for significantly increasing a person's mortality and morbidity. It is a sizable health issue in the UK, with 3 million people expected to be living with all types of diabetes by 2010 (BMA, 2004, page 1). In 2002 approximately 9% of all hospital inpatient expenditure was the consequence of diabetes related illnesses (Department of Health, 2006) and this figure is set to rise as diseases of comfort and their associated health-harming lifestyles become more prevalent in society.

There are two different type of Diabetes; Type I and Type II, with Type II prevalence rates arising from a range of lifestyle related risk factors. National statistics suggest that over 1 million people in the UK may be living with undiagnosed Type II diabetes (BMA, 2004, page 1). The standard current practice for estimating prevalence rates (prevalence is measured as the proportion of a population that are cases at a point in time) of diabetes uses the national "PBS Diabetes population model"; phase II. The model was originally authored by the Public Health Observatory, Brent PCT and the School of Health and Related Research (ScHARR), hence the acronym PBS (YHPHO, 2005). The model "applies age/sex/ethnic group-specific, estimates of diabetes prevalence rates derived from epidemiological population studies, to 2001 Census resident populations" (YHPHO, 2005, online). In doing so it produces estimates of total expected cases of diabetes for a number of administrative units, ranging from Primary Care Trust (PCT) boundaries to General Practice (GP) catchments. The prevalence rates used in the model are derived from surveys conducted in Brent, Coventry and Wales.

Analysis of Camden's 2001 Census population using the PBS model predicted that there were 8,627 diabetics living in Camden (4.36% of population). This statistic includes both diagnosed and undiagnosed cases. If these proportions are applied to the population estimate for 2004 (taken from the 2004 mid-year population estimates as provided by the Office of National Statistics (ONS)) the estimated figure rises to 9,464 people with both diagnosed and undiagnosed Type 1 and Type 2 diabetes. This represents a predicted rise of nearly 10% in the incidence (incidence is measured as the number of new cases arising in a given time frame in a specified population group) of diabetics amongst Camden's resident population.

The PBS model has a number of limitations which inhibits its use. Firstly, the calculations are based upon prevalence amongst different ethnic groups and are derived from survey data collected in just two regional centres (Coventry and Brent, where there are large South East Asian communities). Consequently, either under or over estimation is likely to occur because the model assumes prevalence rates in these two centres are representative of the ethnic minority populations across the entire UK. The current method gives an indication that large numbers of Camden's population live with diabetes, but this finding is limited in its actual application to strategic intervention planning because data are not specific enough for small scale planning. Data are only available for the following administrative units: PCTs, general practices or wards.

The only specific local data available to PCTs is extracted from either the Quality or Outcomes Framework (QOF) or from inpatient admissions databases, known as the Hospital Episode Statistics (HES). The QOF is the annual reward and incentive programme detailing general practice achievement results, whereby practices are awarded points (leading to financial rewards) according to how well they manage chronic (long term) diseases such as asthma or diabetes (QOF, 2007). The QOF reports disease prevalence rates which PCTs use to determine the level of financial rewards to be given to each practice. The prevalence rates available to the PCTs are restricted to aggregate estimates at the scale of each individual general practice. These aggregate rates are difficult to apply to the diverse populations registered by GPs at the neighbourhood level and once again lack specificity of local scale interventions, not based around a general practice. Hence, there is a requirement to develop alternative measures of diabetes risk for informing strategic commissioning practices and health promotion activities at local scales for local areas.

This case study is compliant with objective one and two of the overall research objectives (chapter 4), which are interested in developing a geodemographic framework and applying it to real world scenarios. It sets out to explore and determine the usefulness of national datasets to predict local differences in health. At the same time, the case study will locate different sub groups of populations in order to determine the magnitude of variation in expected diabetes need. Hospital Episode Statistics will be used to predict levels of local hospital usage, and by corollary local levels of diabetes need, by using them alongside a national geodemographic classification.

7.1.2 Datasets and method

Of the range of datasets identified in Chapter 5, only a small number were of use in completing this case study. The ones used are listed in Table 20. They are useful for predicting diabetes need for a number of reasons. The hospital episodes database contains records of patients admitted to hospital, which include patients who have serious or poorly managed diabetes and require an overnight stay. The postcode data of the patient enables the patients to be assigned a geodemographic Type which is subsequently used to predict local neighbourhood need using the spatial location of the Types taken from Codepoint or Gridlink.

Dataset Name	Description	Date
National Hospital Episode Statistics (HES)	Patient record of all episodes related to admission to NHS hospitals in England	1st January 2000 and 31st December 2002
Mosaic Neighbourhood Typology	Commercial geodemographic typology Mosaic produced by Experian, Nottingham, UK. As previously noted, the typology clusters together over 400 different variables, according to similarity and a proximity measure (distance from coast), in order to enable the classification of 1.7 million UK postcodes into 61 different Types.	November 2003
CodePoint/Gridlink	Ordnance Survey product providing spatial coordinated for all postcode unit polygons	2003
General Practices	List of general practices and their associated postcode unit	November 2003
Pharmacies	List of pharmacies and their associated postcode unit	November 2003
Patient list	List of Camden registered, resident patients per general practice, recording age, sex and postcode	November 2005

Table 20: Datasets used in diabetes risk prediction

The Hospital Episode Statistics (HES) used in this study were produced from the Department of Health's national data warehouse for England and record all health episodes relating to patients admitted to NHS hospitals in England. They record information about patient age, sex, postcode, diagnosis and length of stay in hospital. The HES extract used in this analysis included 17 million de-duplicated records of elective, emergency and maternity admissions. This equates to one record for each hospital admission in England, for a three year period, between 1st January 2000 and 31st December 2002. Of the 16,923,845 records 16,599,797 were successfully linked to Mosaic using the postcode. The records that could not be linked were because either the postcode was missing, the disease code was missing, there was no record of age or that the postcode belonged to an unclassified Mosaic Type (Webber, 2004). The data records were de-duplicated to enable the development of risk likelihood statistics which requires unique patient records. If a patient was admitted to hospital multiple times for diabetes then for predicting risk they would only need to be counted once (not for each time they visited hospital).

If the duplicates were not removed the results would be biased towards individuals admitted more than once – which would not indicate risk likelihood but actual usage statistics. Although it could be said that duplicate admissions information could be useful, this is because they would identify people with poorly managed or serious diabetes. For the purposes of this research, because it is investigating proxy indicators for prevalence and risk likelihood, the duplicates were removed. An area of further work would be to explore these duplicate records. This would enable an assessment of neighbourhoods with poorly managed diabetes, a useful indicator for social marketing and interventions around patient management of their diabetes.

The records from the Hospital Episode Statistics used in the case study were supplied to this research project by Experian, Nottingham, via Mr Richard Webber. The original data processing: de-duplicating and cross-referencing with Mosaic, was carried out by the Medical Statistics Unit at Queen Mary Hospital which is part of Imperial College London and was originally commissioned by the consultancy Dr Foster Ltd.

Each patient episode was categorised according to associated diagnosis which were then grouped together according to the International Classification of Diseases (ICD). So for example all episodes classified as diabetes were grouped together as "ICD E10-E14" (WHO, 2007), whereby E10 to E14 correspond to diabetes diagnoses. Using the unit postcode as the key identifier of each patient episode, the geodemographic classification was then appended to the database of hospital episodes (see Figure 5.6, research method). Counts for each neighbourhood Type were cross tabulated for the ICD diagnoses groups and used to calculate a national dataset of total admissions for different neighbourhood Types. It was not necessary to apply a weight to the subsequent risk indices because the original dataset was large enough for the counts to produce robust measures. These standardised risk admission scores were then applied to a number of different administrative units in order to predict likelihood of risk to diabetes and by corollary the patient need of diabetes services which are all useful when considering where to target interventions. The risk scores for each neighbourhood Type or Group were calculated following the method outlined in the research framework, section 5.2 and as applied to the case studies in the previous chapter.

7.1.3 Results and analysis

Using the method outlined in the previous chapters, section 5.2, index risk scores for hospital admissions were calculated for each of the 61 neighbourhood Types and 11 Groups. The average number of diabetes episodes in a neighbourhood was compared to the average number of diabetes related episodes in England, to create the risk indices (where an index value of 100 is the equivalent of the national average). This was calculated for all diabetes related episodes and diabetes Type II episodes (usually triggered by poor lifestyle). A variety of methods were employed to predict diabetes risk for both local neighbourhoods and General Practices which are discussed in the following sections.

Table 21 highlights the risk index scores for all diabetes as calculated for the neighbourhood Groups in England. The risk index scores calculated using data from 2000-2002 are listed in the column on the far right of the table. To summarise, 1 in every 20 hospital admissions in England is related to diabetes conditions. There is considerable variation in the resultant scores, with a minimum neighbourhood risk score of 68 and a maximum of 149. The Group most likely to suffer from diabetes related illness fall within the cluster Group known as Twilight Subsistence (Group I). People living in this sub-population group have the characteristics of being old (over 65 years of age) with low incomes and are likely to be living in social housing. People living in this neighbourhood Group are responsible for 1 in 12 of all diabetes admissions in England. For Group D, a neighbourhood group predominated by South Asians, a staggering 18% of all hospital admissions for the Group are the result of diabetes. Moreover, one sixth of all diabetes admissions in England could be attributed to people living in this Group. Even these basic statistics can be useful for social marketers, whereby campaigns are targeted at population sub-groups who appear to badly manage their diabetes or have a serious case of the disease.

Group	Number of Diabetes Admissions	% Target	Total Admissions For Type	% Base	Risk Index Score
a	49289	6%	1296015	8%	80
b	42427	5%	1308790	8%	68
c	120201	15%	2597926	16%	97
d	143032	18%	2958969	18%	102
e	33234	4%	801349	5%	87
f	50334	6%	993078	6%	107
g	76314	10%	1522272	9%	106
h	102249	13%	2029290	12%	106
i	63234	8%	890754	5%	149
j	78946	10%	1518838	9%	109
k	29670	4%	691391	4%	90

Table 21: Index of Diabetes admissions for different neighbourhood Groups

To estimate the need for local diabetes services and predict the risk likelihood of residents to diabetes related illnesses, local need for each neighbourhood Type was calculated. The national index scores for each Type were linked to both the postcodes in Camden using the patient registers maintained by each general practice. This facilitated a variety of analysis to identify residential postcodes where residents were most likely to be admitted to hospital for diabetes related illness, all types and diabetes type II. The analysis is described in the following sections, exploring the link between the measures and scale.

7.2 Scaling the indicators

In the literature review the four key principles of social marketing were identified as product, price, promotion and place. With a thorough understanding of the place aspect of social marketing it is possible to develop the remaining principles within a campaign or public health intervention. But as de Smith and colleagues note (2007), 'on its own spatial location is not interesting. The power of location comes not from location itself, but from the linkages or relationships that it establishes – from relative positions rather than absolute ones' (de Smith et al., 2007, online). It is the attributes and features associated with a location that make it a place that is interesting for research.

Thus far, geodemographic classifications have enabled social scale to be explored for different population sub-groups according to a diverse array of socio-economic and lifestyle variables which are then related to a specific place, in this instance a neighbourhood postcode unit. The research has looked at ascribing neighbourhood characteristics to small local neighbourhoods in accordance with the unit postcode. Within the public sector, this scale of analysis whilst preferable is unusual. This is because without the application of a neighbourhood typology it would be difficult to assign population characteristics to the local scale without making broad assumptions which ignore the heterogeneity of an area and make assumptions based on homogeneity.

Considering the importance of both the social and spatial scale in parallel, the following section explores how the framework and resultant indicators can be scaled to different geographical extents which have strategic importance for health policy decision-makers and strategists. In this next section diabetes is predicted for three different scales, which are organised into a nested hierarchy; neighbourhoods, general practices and primary care trusts, illustrated in diagram, Figure 42. This hierarchy is comprised of two different scales, the formal scale which corresponds to the organisational scale of the NHS and the functional scale which represents the social scale of the neighbourhood. A greater discussion on the implications of functional versus formal scale is returned to in Chapter 8.

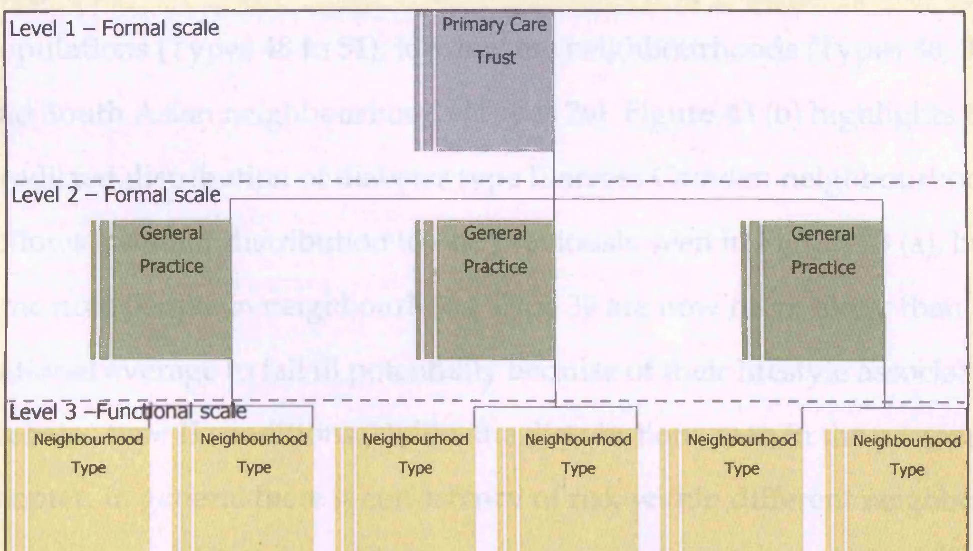
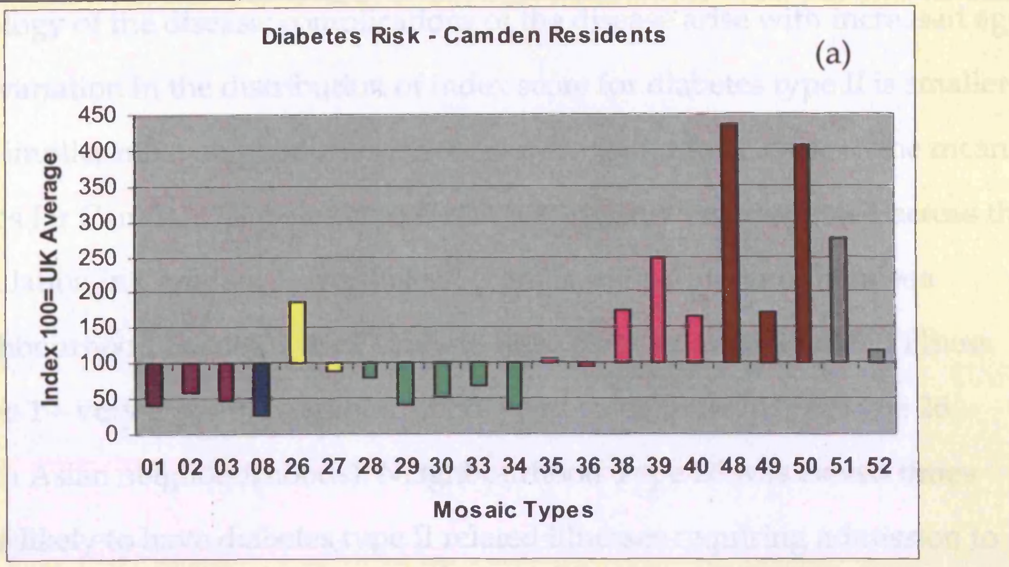


Figure 42: Functional and formal scales used to predict diabetes need

7.2.1 Predictions at the scale of the neighbourhood

The first scale explored is the neighbourhood. In this section as previously noted, the concept of a neighbourhood relates to a unit postcode. The neighbourhood risk index scores were linked to the local postcode code directory for the London borough of Camden as described in section 7.1.2, supporting the development of a local diabetes risk profile for application to local communities. Diabetes risk was then associated with the residents in each postcode neighbourhood, using the same method described in section 5.2. Figure 43 shows all diabetes risk (type 1 and 2 combined) for Camden residents, based upon the neighbourhood Types. On both graphs the value of 100 acts a benchmark, indicating the national average. For neighbourhood Types with scores in excess of 100 the likelihood of a resident presenting to hospital with diabetes is greater than the national average.

Figure 43 (a) highlights the predicted distribution of all diabetes across neighbourhoods; note the slightly different scales of the two graphs. Neighbourhood Types with the greatest risk are represented by elderly populations (Types 48 to 51), low income neighbourhoods (Types 38, 39, 40) and South Asian neighbourhoods (Types 26). Figure 43 (b) highlights the predicted distribution of diabetes type II across Camden neighbourhoods. It follows a similar distribution to one previously seen in Figure 43 (a), but this time note people in neighbourhood Type 36 are now more likely than the national average to fall ill potentially because of their lifestyle associated with diabetes type II conditions. Unlike the distributions seen in the previous chapter, in general there is consistency of risk within different neighbourhood groups (bars with the same colour on the graph) and larger differences between groups.



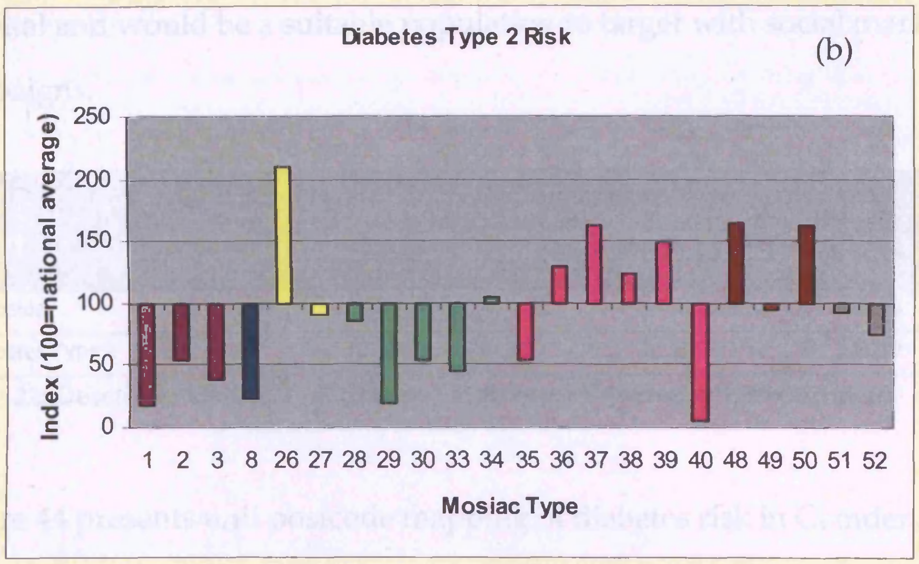


Figure 43: Risk index scores for Camden neighbourhoods: (a) all diabetes, (b) type II diabetes

The standard deviation of the index scores was used to indicate the degree of variation in the risk distributions across different neighbourhoods in Camden. It measured the extent to which diabetes risk varied about its mean for different population Types. As shown in Table 22, the standard deviation of diabetes risk for all types of diabetes is 117. The biggest difference is between Type 08 (young families in new housing) and Type 48 (old people in flats). Older people in flats are sixteen times more likely than young families to have serious diabetes related illness. This indicates that older people are more vulnerable to being admitted to hospital with diabetes, reflecting the aetiology of the disease; complications of the disease arise with increased age. The variation in the distribution of index score for diabetes type II is smaller. This implies most neighbourhood scores were found to be close to the mean scores for Camden. Diabetes type II risk was slightly less dispersed across the population in Camden. Nevertheless, there is still a difference between neighbourhood residents least likely to have diabetes type II related illness (Type 1 – very affluent neighbourhoods) and those most likely (Type 26 – South Asian neighbourhoods). Neighbourhood Type 26 was eleven times more likely to have diabetes type II related illnesses requiring admission to hospital and would be a suitable population to target with social marketing campaigns.

	Number of neighbourhoods	Minimum index score	Maximum index score	Mean index score	Standard Deviation
Diabetes	20	26	436	139.19	117.45
Diabetes Type 2	20	18	209	87.95	52.810

Table 22: Descriptor statistics for diabetes indices for Camden neighbourhoods

Figure 44 presents unit postcode mapping of diabetes risk in Camden. The maps were classified into 5 equal groups, reflecting dispersion around the national average (index score = 100). Neighbourhoods with index scores 0 to 50 were classified as considerably below average, 50 to 100 were below average, 100 to 150 were average, 150 to 200 were above average and greater than 200 were considerably above average. The maps have been coloured using a technique known as hot-cold mapping, moving from a pale colour to a dark colour palette. A red – pale yellow colour gradient was applied so neighbourhoods with the greatest risk of diabetes admissions were coloured in red, and those with the lowest risk in pale yellow.

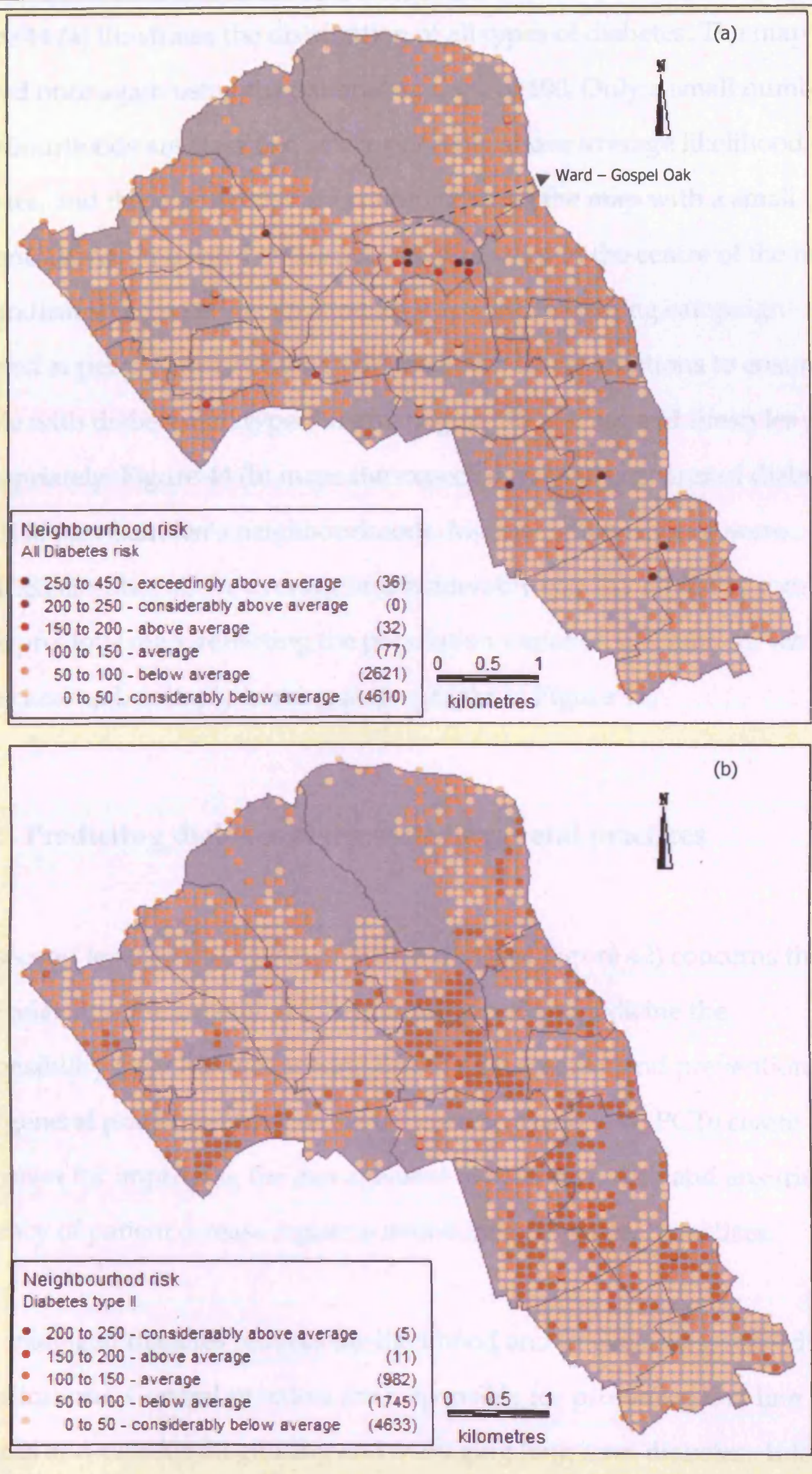


Figure 44: Postcode risk index maps for Camden neighbourhoods: (a) all diabetes, (b) type II diabetes

Figure 44 (a) illustrates the distribution of all types of diabetes. The map was centred once again using the national average of 100. Only a small number of neighbourhoods are classified as considerably above average likelihood of diabetes, and these postcodes are scattered across the map with a small concentration in the ward of Gospel Oak (the ward in the centre of the map). This indicates the potential location for a specific marketing campaign targeted at people already living with diabetes, ie interventions to ensure people with diabetes (all types) are managing the disease and lifestyles appropriately. Figure 44 (b) maps the expected risk index scores of diabetes type II across Camden's neighbourhoods. More neighbourhoods were classified as either above average or considerably above average as compared to the previous map, reflecting the population variation of Camden, which is not picked up by simply looking at the graphs in Figure 43.

7.2.2 Predicting diabetes at the scale of general practices

The second level of scale we are interested in (see Figure 42) concerns the scale related to the general practice. In primary care medicine the responsibility for initiatives around health management and prevention lies with general practices. Many of the GP contracts held with PCTs create incentives for improving the management of diabetic health and ensuring accuracy of patient disease registers maintained by general practices.

Well managed diabetes reduces the likelihood and impact of associated complications. General practices are responsible for providing first line services and care for diagnosing and managing long term diabetes. If local demand can be estimated then adequate and equitable service commissioning will contribute to the establishment of viable, effective and efficient social marketing solutions. If the estimated likelihood of demand is calculated for general practices in Camden, health practioners are better placed to understand the potential demand for diabetic services and the risk of undiagnosed diabetes.

Composite scores for general practices in Camden were created for all diabetes and diabetes type II related risk. An honorary contract with Camden PCT enabled access to patient lists for general practices in Camden; which is the inventory of individual patients registered to general practices identifying basic individual characteristics; age, date of birth, address and postcodes. Using this patient list for each general practice the proportion of patients in each neighbourhood Type for each practice was calculated. These proportions were then multiplied by the corresponding diabetes risk index score, as used in the previous section. An example is illustrated for the general practice with the code "F83003" (the unique identifier of that practice). This general practice, in the north of the borough, in the ward of Hampstead Town has 4872 patients from neighbourhood Type 1 registered; representing 70% of its total patients. This is then multiplied by the associated index score for diabetes and in this instance diabetes type II of that neighbourhood which is 18 to give a result of 12.3 [(4872/7076)*18]. The process is repeated for each neighbourhood. The results for each neighbourhood Type for each practice were then summed in order to produce a composite risk score of expected diabetes demand for each of Camden's general practices.

Once again the benchmark of 100 was set as the national average. All of the practices in Camden were below this value. Thus, the case mix (different types of cases that results from different types of patients) of patients registered at Camden general practices ensured none of them exhibited above average likelihood of all diabetes admissions. This distribution is highlighted in Figure 45 (a). If diabetes type II risk is predicted only, Figure 45 (b), the composite risk scores increase for all general practices. Population sub-groups in Camden are more at risk of lifestyle related diabetes type II. There is one particular practice highlighted in Figure 45 (b) which is located in the centre of the borough, which has a considerably above average risk that their registered patients are likely to develop lifestyle induced diabetes type II. This pattern for diabetes type II can primarily be described by looking at neighbourhood Type 36; ethnically diverse neighbourhoods, who are often reliant on benefits and appear to have above average likelihood of lifestyle related diabetes type II (and not all diabetes). People living in this neighbourhood comprise almost one quarter of Camden's resident population. Of the forty six practices in Camden, nine have in excess of 48% of registered patients from this neighbourhood Type alone.

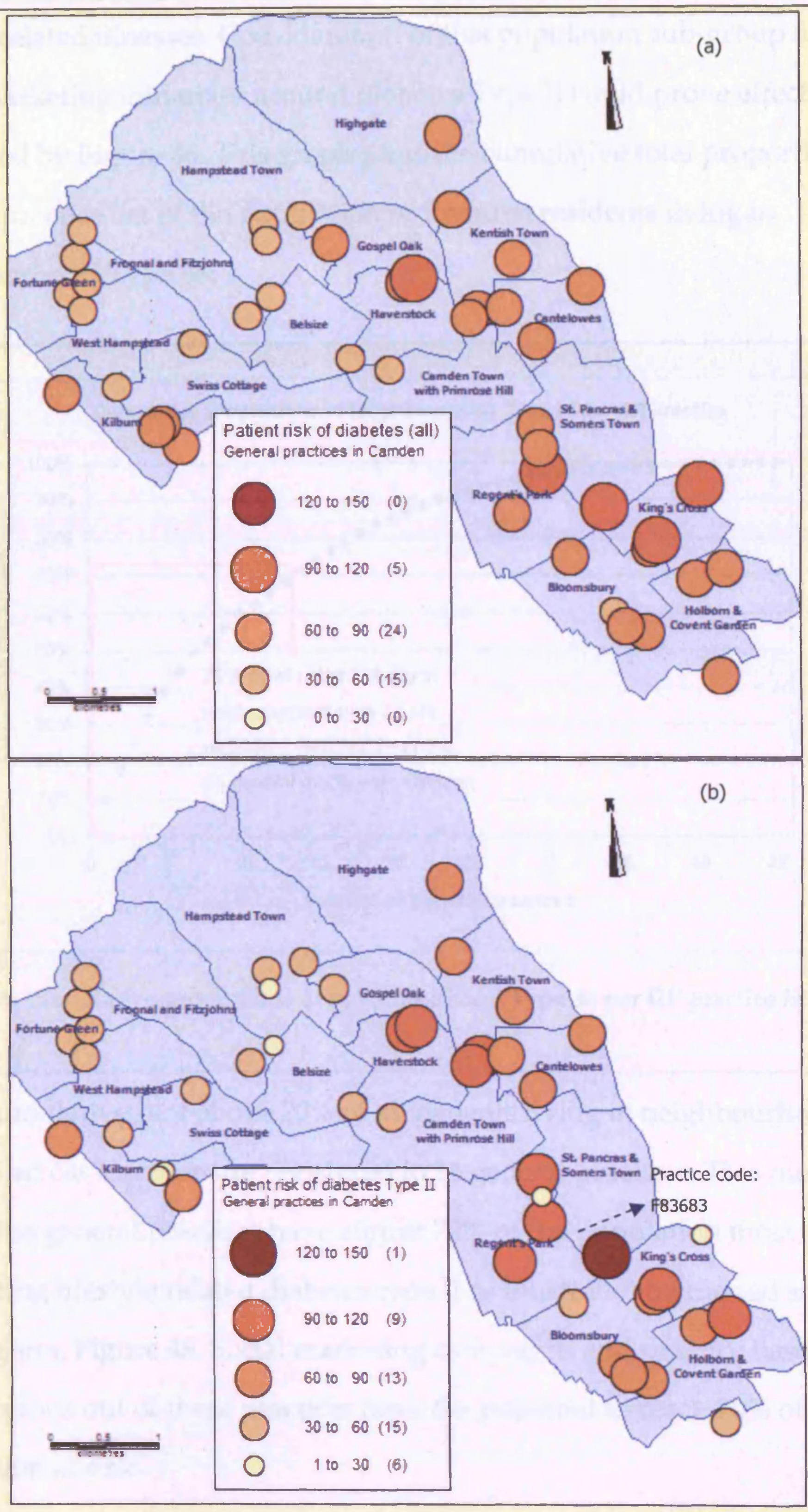


Figure 45: Composite risk index maps for Camden general practices: (a) all diabetes, (b) type II diabetes

The diabetes type II risk profiles indicated people living in Type 36 neighbourhoods are much more likely to be admitted to hospital for diabetes

Type II related illnesses. Consideration of this population sub-group for social marketing initiatives around diabetes Type II could prove effective as illustrated by Figure 46. This graph plots the cumulative total proportions per general practice list of the population of Camden residents living in neighbourhood Type 36.

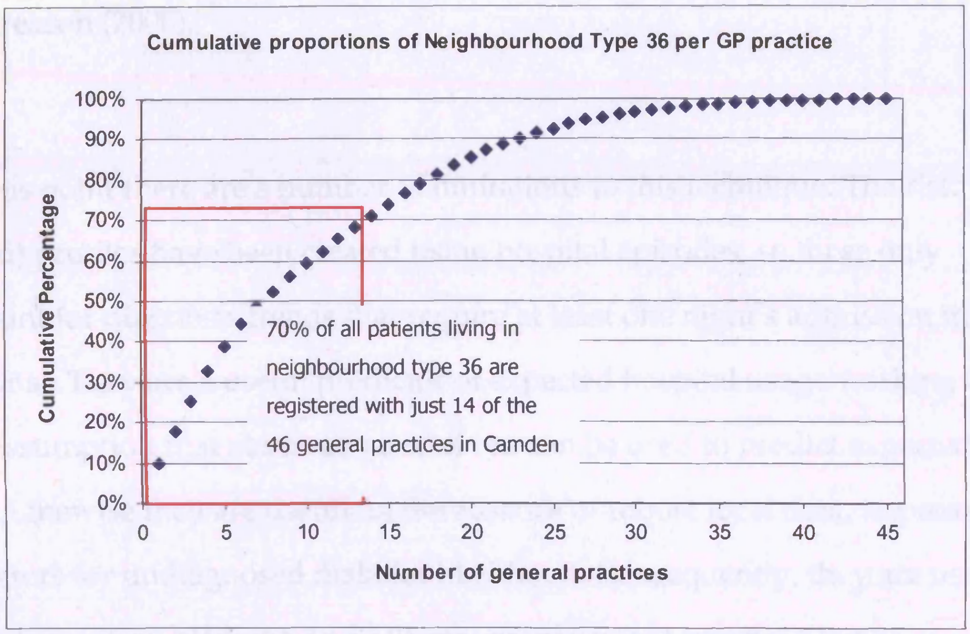


Figure 46: Cumulative proportions of neighbourhood Type 36 per GP practice in Camden

In the cumulative plot above 70% of all patients living in neighbourhood Type 36 across Camden are registered to 14 general practices. This means that 30% of the general practices have almost 70% of the population most at risk of developing lifestyle related diabetes type II as illustrated by the red square in the diagram, Figure 46. Social marketing campaigns and practice based interventions out of these practices have the potential to reach 70% of population at risk.

This more targeted approach would reach the population at greatest risk and be more efficient in terms of resources and cost as opposed to taking a blanket approach of the whole population in Camden. Although the main consideration is the ethical implications relating to fairness and equity, what would happen to the 30% of neighbourhood Type 36 who are not registered at these practices? Indeed it could even be said that the remaining 30% are the most in need because they are registered at practices where they are the minority registered population group, and services at these practices are already less likely to be targeting their needs. The ethics of social marketing is beyond the scope of this research and are a subject of recent research by Andreason (2000).

At this point there are a number of limitations to this technique. The risk (need) profiles have been created using hospital episodes, so these only account for diagnosis trends that require at least one night's admission into a hospital. They are a useful predictor of expected hospital usage working on the assumption that observed admissions can be used to predict expected risk. Likewise they are useful, in the absence of robust local data, as proxy indictors for undiagnosed diabetes likelihood. Consequently, they are useful for intervention planning and self-care management planning but as a predictor for actual disease prevalence in a population they have limited application.

7.2.3 Predicting diabetes risk at the scale of the Primary Care Trust (PCT)

In the same way that composite risk scores of diabetes were created for GPs in Camden, this technique was then applied to the final scale in the hierarchy, level 1 (Figure 42) to estimate risk for the administrative boundary comprising the area of responsibility for Camden. This fits with the scale of decision making. Composite risk scores were calculated for each PCT in London by utilising the same method that was applied to create the diabetes practice profiles (section 7.2.2). These scores were then mapped in order to show the distributed risk scores expected for all types of diabetes admissions for PCT boundaries, as shown in Figure 47.

In Figure 47 the composite risk scores for PCTs have been mapped, identifying PCTs with the greatest risk of patients being admitted to hospital coloured in red (score over 90) and scores below the national average in light blue. The standard deviation for the PCTs all diabetes distribution was calculated as 13.07, with a mean index score of 83.7; the distribution is very narrow and centred on the mean, which is below the national average. If the standard deviation was larger then PCTs would have more variation in the distribution of diabetes risk. As it is, index scores for PCT administrative boundaries in London do not differ greatly from each other; 19 of the 32 PCTs have a composite all diabetes risk score of between 30 to 90.

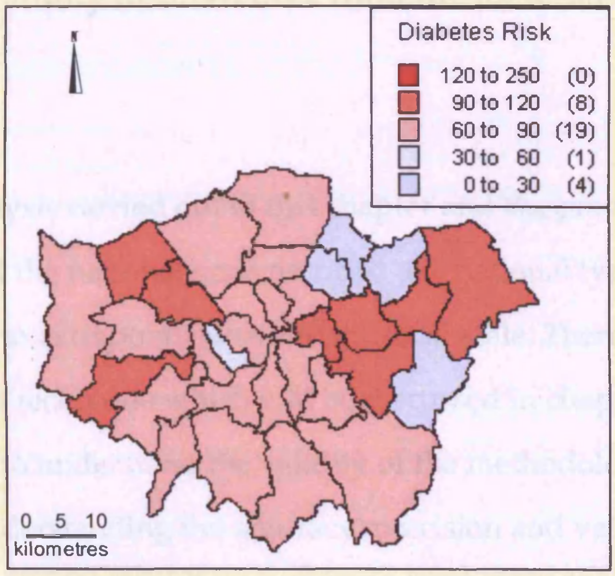


Figure 47: Predicted diabetes admissions for London PCTs

There are two PCTs that have expected admission rates greater than the national average; the London Borough of Newham together with that of Barking and Dagenham; although greater need is indicated for PCTs in the east and northwest London. Meanwhile the lowest rates of diabetes admission are expected for Kensington and Chelsea. Camden ranks in the middle alongside Sutton and Merton, Wandsworth, and Kingston. At this scale it is difficult to distinguish actual diabetes risk to populations, because they have more or less similar expected risk. So whilst the indictors can be scaled up, local decision-making at this scale would be inappropriate, because local distributions are not identified and real need is hidden; this is the inevitable result of the influence of scale which subjects the results to limitations of the modifiable areal unit problem and the ecological fallacy (Openshaw, 1984; Harris et al., 2005), an issue returned to in Chapter 8. To understand broader scales of need, further investigation is required. This could be fulfilled by using local indicators of spatial association on local data to determine geographically significant clustering, highlighting diabetes need more according to more appropriate analysis units.

7.2.4 Testing validity of inferences from the national to the local level

So far all the analysis carried out in this chapter and the previous one has harnessed data at the national scale ascribed to a national typology of neighbourhoods to extrapolate down to the local scale. There are a number of limitations to this technique which will be discussed in chapter 8, but this section attempts to understand the validity of the methodology. In social measurement understanding the accuracy, precision and validity of the metrics is critical. The limitations of using what is essentially a "black box" commercial neighbourhood classification presents challenges when trying to ascertain the validity of the actual classification because the exact input variables, their weightings and the detailed methodology used to create the products remains the intellectual property of these companies, and so precision and validity cannot be determined. Consequently, the challenges of understanding validity and evaluating their uncertainty are then transferred to the health outcome indices which have been developed.

In attempting to understand the accuracy of the developed indices, this next section looks at two methods of exploring and validating the results. The first method predicts local hospital episodes of diabetes from the national dataset and then compares it to the actual number of local hospital admissions by residents of Camden. This is conducted because it will identify how close both the actual and predicted results are. The second method compares a nationally accepted model used to predict diabetes (see section 7.1.1) and compares the results with the health outcomes indicator for diabetes.

7.2.4.1 *Observed versus predicted numbers of diabetes episodes*

Using the proportions of the national population in each neighbourhood Type it was possible to develop local estimates of expected episodes of diabetes related illnesses (a proxy for predicting diabetes risk in the population). The total residential population living in each Camden neighbourhood Type was calculated using the number of residents per postcode unit (derived from the data contained in the geodemographic dataset Mosaic). The Camden population totals per neighbourhood Type were then multiplied by the national percentage of diabetes admissions for each neighbourhood Type. For example 58,665 people live in neighbourhood Type 26 (25% of Camden's total population). If the proportion of all people admitted for all diabetes per Type (nationally) is multiplied by 58665, it could be predicted that 84 people would be expected to be admitted to hospital (during the time frame between Jan 2000 to Dec 2002). The resulting values were a predictor of the expected number of individual local diabetes admission episodes given national rates of hospital usage.

The actual numbers of hospital episodes per neighbourhood Type for the time frame of the study were extracted from the hospital episodes database. They were then evaluated against the expected number of episodes highlighted in the graph in Figure 48. The actual number of hospital episodes are coloured in blue and the expected number of episodes in green. On observation of the graph there appears to be broad similarity between the observed (actual number of local admissions) and the expected number of admissions predicted from the national dataset, leading one to conclude that national proportions of all diabetes related hospital episodes for neighbourhoods can be accurately used to extrapolate to the neighbourhood scale (at least for Camden, London).

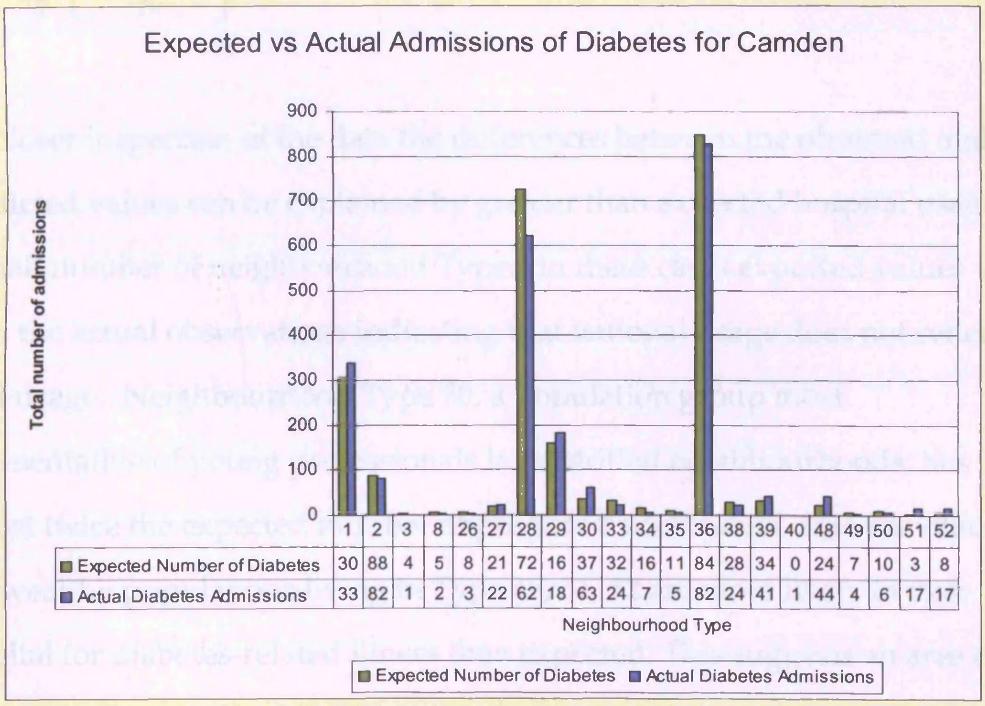


Figure 48: Actual versus estimated numbers of hospital episodes by neighbourhood Type over a 3 year period

The goodness of fit between the actual admissions and the predicted admissions can be ascertained using the Chi squared statistic, see equation below. This is a statistic used to identify whether the observed similarities between national predictions (expected E_i) and local observations (observed

O_i) are statistically significant or the result of random chance. Thus, the null hypothesis (Ho) is stated as: there were no significant differences between the observed and the predicted (expected) results. The resultant Chi squared (X^2) statistic was 95.01 with 20 degrees of freedom (n=21, degrees of freedom = n-1). Using a significance level of p>0.01, the critical Chi squared value with 20 degrees of freedom signals the rejection of Ho at values above 37.57. The rejection of the null hypothesis leads to the conclusion that local patterns of hospital episodes do differ significantly from national predictions of local episodes at the level of the geodemographic Type, the reason for which is explored next.

$$\chi_2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)2}{E_i} \right] \qquad \text{Equation 3}$$

On closer inspection of the data the differences between the observed and the predicted values can be explained by greater than expected hospital usage by a small number of neighbourhood Types, in these cases expected values vary from the actual observations indicating that national usage does not reflect local usage. Neighbourhood Type 30, a population group most representative of young professionals in gentrified neighbourhoods, has almost twice the expected number of predicted admissions. Also the older, less wealthy populations living in Type 48, 51, 52 are more likely to visit hospital for diabetes-related illness than expected. This suggests an area of further work at the local level to investigate why these population Types are have poorly managed diabetes in Camden and differ significantly from national predictors of diabetes admissions.

Looking more closely at the data one of the largest differences between the observed and the expected values, with more than 14 actual episodes than expected was observed for Neighbourhood Type 51. This Neighbourhood

Type was removed from the dataset and the Chi squared statistic recalculated. It was removed because the very small expected value (3) was exerting too much influence on the test statistic calculation. The null hypothesis (Ho) was restated; there were no significant differences between the observed and the predicted values of hospital episodes. The new Chi squared value was 29.68, with 19 degrees of freedom. Using the same significance level of p>0.01 the critical value of the Chi squared statistic was 36.1. In this instance the null hypothesis is below the critical value which means it can be accepted. There is less than a 1 in 100 chance that the distribution is the result of random chance and it is therefore probable that no significant differences exists between the results of predicting neighbourhood diabetes need from national data and actual observable need extracted from local administrative records. This highlights the sensitivity of statistical analysis to small numbers in datasets.

7.2.4.2 Comparison with a standard national model of diabetes prevalence:

Composite risk indicators using geodemographics and operational health to predict expected health outcomes are a new concept within the field of Public Health. In the previous section the validity of the measures was explored by comparing actual observations against predicted estimates of need to determine the goodness of fit between the two. In this section another method for exploring the precision and validity is presented. Rankings of PCTs were carried out using both a known and accepted model for predicting diabetes prevalence at the scale of the PCTs and the geodemographic framework presented in this thesis (Chapter 5). The nationwide standard for predicting diabetes need is the PBS model[1], which measures diabetes prevalence for PCTs, general practices and administrative wards.

Table 23 lists the PCTs and their associated diabetes risk scores calculated using both models; geodemographics and the conventional prevalence rates predicted using the national PBS model. Strictly speaking the two models are not comparing exactly the same outcome; the national model predicts diagnosed and undiagnosed prevalence whereas the neighbourhood indicators predict demand for hospitalisation caused by diabetes (which is used as a proxy for diabetes risk). This research makes the assumption that neighbourhoods Types with high propensities for hospital usage relating to diabetes will have risk rates of diabetes in broadly the same proportions.

With the PBS model the PCT with the highest population prevalence of 6% for diabetes is listed as Brent. The geodemographic model predicted diabetes risk for Brent as lower than the national average, but using the PBS model its prevalence rates are much higher. This result is likely to be an artefact of the computation within the PBS model because it was built using survey data from this London Borough (see 7.1.1). Furthermore, whist the geodemographic Types do cluster similar people together which includes clustering age groups, the index scores are not standardised by age, so this may contribute to the differences between the two models.

[1] The acronym PBS is derived from the initials of the organisations that authored the tool; Public Health Observatory, Brent PCT, School of Health and Related Research (ScHARR),

PCT Code	PCT Label	Composite Diabetes Risk Score	Diabetes Prevalence (PCT) (2001 PBS prevalence model)
5K5	Brent PCT	93	6.00%
5C5	Newham PCT	113	5.64%
5C8	Redbridge PCT	95	5.23%
5HX	Ealing PCT	87	5.13%
5C3	City and Hackney PCT	86	5.10%
5C4	Tower Hamlets PCT	96	5.09%
5K6	Harrow PCT	91	5.04%
5C6	Waltham Forest PCT	93	5.01%
5HY	Hounslow PCT	91	4.90%
5C9	Haringey Teaching PCT	80	4.80%
5LC	Westminster PCT	61	4.67%
5C1	Enfield PCT	88	4.65%
5LE	Southwark PCT	84	4.64%
5C2	Barking and Dagenham PCT	110	4.62%
5A8	Greenwich PCT	95	4.61%
5K8	Islington PCT	81	4.58%
5K9	Croydon PCT	84	4.57%
5LD	Lambeth PCT	77	4.53%
5LF	Lewisham PCT	88	4.48%
5K7	Camden PCT	70	4.35%
5A9	Barnet PCT	74	4.35%
5AT	Hillingdon PCT	88	4.24%
5A4	Havering PCT	94	4.15%
5LA	Kensington and Chelsea PCT	59	4.10%
5H1	Hammersmith and Fulham PCT	68	4.09%
5AX	Bexley Care Trust	90	3.99%
5A7	Bromley PCT	81	3.99%
5M7	Sutton and Merton PCT	78	3.86%
5LG	Wandsworth PCT	66	3.77%
5A5	Kingston PCT	68	3.51%
5M6	Richmond and Twickenham PCT	65	3.34%

Table 23: London PCT composite risk scores for all types of diabetes, ranked by prevalence as predicted by the PBS model

If composite geodemographic scores are a viable alternative to other traditional models of calculating diabetes prevalence, then it follows that there should be a degree of correlation in the ranking between the two techniques. This is because both models predict diabetes need in the population albeit using different base data and, processes. If the results differ widely it is fair to assume that inaccuracies have been introduced into the modelling process that would need to be addressed in future research. A number of assumptions have to be made prior to comparing the models which must be mentioned. Firstly that the PBS model is a reflection of diabetes prevalence and if the geodemographics framework is a suitable alternative then there will be correlation between the two modelling results. Secondly, in the case of the geodemographics model the results assume hospital admissions for diabetes are a suitable proxy indicator for actual disease prevalence in the population.

To determine the extent of correlation between the two model based methods that predict diabetes, a bivariate Spearman's rank correlation test was conducted between the prevalence rates of diabetes predicted by the PBS model and the scores returned by the geodemographic model. Spearman's rank test was used to discover the strength of the association between the two models for two reasons. First, because for each PCT there was a pair of data ranked in order (i.e. they are ordinal) for predicted prevalence and secondly because the data were not normally distributed (which rules out using the Pearson's correlation coefficient). The Chi-squared test was an inappropriate test for this example because unlike in the previous section, this example did not have data for both the observed and expected outcomes. To conduct a Spearman's rank test there are two pre-requisites; the data should be on an ordinal scale and there must be at least 5 pairs of data.

The test looks at determining whether the two prevalence scores of diabetes covary; if both models are returning similar rankings for diabetes prevalence for PCTs then the two variables would exhibit covariance. If the PBS model predicted high expected prevalence for a PCT one would expect the geodemographic model to do the same (if it is a valid model). In this example the null hypothesis being tested states that the rank order of the PBS model does not co-vary with the rank ordering produced by the geodemographic model of diabetes prevalence. The formula for testing the Spearman's rank correlation coefficient was outlined previously. The correlation coefficient calculates differences between the ranks of the prevalence scores of the two models.

For this test, the significance was tested at the 0.01 significance level with 31 data pairs (ie the number of PCTs in London (n=31)). A two-tailed test was used because the relationship between the sets of prevalence scores was unknown. At the significance level of 0.01 with 29 degrees of freedom (d.f. = n-2) the critical value is 0.312. For the test to be significant the observed correlation coefficient (r_s) must equal or exceed this critical value. The results are highlighted in Table 24.

Spearman's rank correlation for ordinal data		Diabetes prevalence predicted by geodemographics	Diabetes prevalence predicted by PBS model
r_s	Diabetes prevalence predicted by geodemographics	1.000	.608
	Diabetes prevalence predicted by PBS model	.608	1.000

Table 24: Table showing two-tailed Spearman's rank correlation of Diabetes index scores and diabetes prevalence rates

The observed correlation coefficient for the two-tailed test is greater than the critical value. It also indicates a positive relationship. The Spearman's rank correlation identified a significant association between the rankings of the two sets of predicted prevalence scores across the 31 London PCTs. There was less than a 0.01 (1%) probability that the correlation coefficient of this size could have occurred by chance. The result reflects a true non-random positive association between the two models. For this case study there is a significant association between a PCTs composite geodemographic risk score for diabetes and the prevalence rates of diabetes as predicted by the conventional PBS diabetes model. Thus predicted distributions of local hospital usage using geodemographics broadly rank PCTs in the same order of lowest to highest risk as the PBS model, and vice versa.

7.2.4.3 Discussion

This case study and all of its components presented an alternative approach to exploring diabetes risk in local populations, which is necessary because of the general lack of good quality local data. The approach assists the acquiring of local geographical and health knowledge based on social similarity. The case study illustrated the possibilities of using alternative methods to enable prediction of diabetes risk for a number of different spatial and administrative scales, ranging from the local neighbourhood level to the larger administrative boundaries of Primary Care Trusts. By calculating the diabetes risk for neighbourhoods Types at the national scale and extrapolating down to the local scale, these indicators can then be manipulated for the existing (if not always appropriate, see Chapter 8) scales of analysis required by policy makers for decision-making.

In section 7.2 national rates of diabetes episodes were used to create risk scores of likely diabetes need, which were then applied to local neighbourhoods (unit postcodes). This enabled neighbourhoods with elderly populations or mixed ethnicity (predominantly South Asians) to be identified as those most at risk. A secondary use of these indicators is to predict local need, based according to hospital usage. Neighbourhood profiles of local risk provide an enriched picture of community health needs. National patterns of hospital episodes were a successful predictor of actual usage and need in the case of Camden.

The outputs from 7.2 were used by the Camden PCT to inform a diabetes social marketing campaign and health promotion events. As noted in the literature review there are four main principles of social marketing; product, price, place and promotion. For the intervention campaign the analysis identified the product (a healthy long life), the price (shorter life expectancy and extended morbidity resulting from diabetes), the place (neighbourhood locations with above average likelihood of diabetes type II) and the promotion (use of health bus to carry out street interventions to identify individual risk factors for the disease). The community profiles identified the most likely population sub-groups that would be seen at the different places and ensured suitable literature and staff were available.

The validity of the method was tested to determine if the adoption of such an approach is useful for public health practitioners. Comparisons were made by testing actual hospital usage statistics for hospitals in Camden against predicted levels of usage. A chi-squared test showed that national rates could convincingly predict local usage in Camden only if neighbourhoods with very small numbers (below three) were removed from the test data. The second test compared the rankings of composite diabetes risk scores for PCTs in London against a nationally approved and widely used prevalence model. Again, the rankings of PCTs for both models did not deviate significantly and were not the result of random chance.

The risk profiles for London PCTs developed in section 7.2.3 produced the same rankings of PCTs based on their diabetes need as the prevalence rates of PCTs predicted by the national PBS model, highlighting that the results were likely not a result of random chance. They are a valid and viable alternative for predicting diabetes hospital usage for PCTs. The application of the risk index scores to different spatial units whilst limited by the complexities of ecological fallacy do enable the measures to be scaled. The extent of the disproportionalities resulting from diabetes risk variation across the local Camden neighbourhoods is not as evident from the PCT profiles (section 7.2.3) produced by either this geodemographic model or the conventional model that predicts diabetes using census data and population estimates (PBS model), this is because the large aspatial units of PCTs is not sufficient at identifying social polarities and heterogeneous neighbourhoods. Indeed, to reiterate the argument that policy should not be decided on data aggregated at such large scales. Indeed the PCT profiles have little or no use for health interventions and social marketing campaigns.

Whilst this technique is not an absolute measure of either service demand or need or actual disease prevalence, it indicates potential hospital use and in the absence of any other robust, reliable data, it provides a useful tool for synthetically estimating diabetes risk for segmenting and targeting population sub-groups as it enables one to recognize the important elements of social marketing; product, price, place and promotion. It is a feasible data driven application of geodemographics for use within the social marketing arena. A wider consideration about the suitability of geodemographic indicators for social marketing will be made in the discussion Chapter 8.

7.3 Practice based profiling

Demand is placed upon on local PCTs to commission equitable services that ensure access for those who need them. "Equity of care is a founding principle of the NHS, but healthcare in London is not always equitable, either in terms of mental and physical health outcomes, or in terms of the funding and quality of services offered' (Healthcare for London: a Framework for Action, page 4). The primary goals of health inequality initiatives are explicitly related to comparisons of outcomes between population groups and reducing known disparities and identifying new ones. Despite this there is an ever increasing pressure to improve the efficiency of service provision of which social marketing is just one element, whilst maintaining the health needs of the population.

It is essential to understand that different neighbourhoods and communities require different health promotion and screening strategies, and that different general practices will need to provide individual or shared services according to the profile of their patients, in order to maximise the benefits to all. In section 7.1 of this chapter practice profiles were developed for diabetes-related illnesses and the flexibility of the measures was highlighted. The capability of the indices to be scaled and adapted is further substantiated in the next section, which further investigates and substantiates their use in practice profiling of diseases and conditions.

7.3.1 Problem definition

As the focus of the health service increasingly extends beyond the clinical treatment of patients, in an attempt to reduce inequalities and to improve the health of local populations, so preventative campaigns and promotional initiatives assume greater tactical and strategic importance. The focus of targeting extends likewise to the identification of neighbourhoods that have the highest risk of particular diagnoses and general practices whose patient lists have a higher likelihood of particular diagnoses.

Once again Camden, inner London, is used as the test bed for this case study. Camden has a diverse population with extreme polarities of wealth existing between different communities. The application of geodemographic analysis to registered patient data enables both description and identification of patterns of variations of different diseases. It provides added-value information to those engaged in service delivery, planning and promotion and social marketing. This technique permits services to be tailored appropriately to reflect the population diversity and their differing health needs. A detailed comprehension of population diversity will assist PCTs in the creation of opportunities through provision of support and information (social marketing) to make it possible for patients to take informed health decisions and improve service compliance.

Once again the issues of accessibility and data protection issues surround and hinder Camden Primary Care Trust's (PCT) ability to access individual and aggregate data collected by general practices relating to specific diseases. The exploratory data analysis documented in this case study combines commercial marketing techniques for segmenting populations that were explored in the previous case study together with data extracted from the Department of Health hospital episode statistics to investigate differences in admission patterns. Building upon the diabetes case study (section 7.1), the data are applied to the local level. Data were linked to the patient registration database and led to the production of practice profiles of disease likelihood through the creation of proxy indicators of health risk outcomes.

7.3.2 Method and analysis

This case study used the same dataset as the previous one, whereby all admissions were then coded by Mosaic, using the postcode of each patient as the identifier. Overall records were coded into one of the 21 classes set out in Figure 49.

Diagnosis Groups	code
All inpatients	99
All cancers	00
Breast cancer	01
Lung cancer	02
Cervical cancer	03
Bowel cancer	04
Other cancers	05
Aged 65+ Emergency admissions (exc mental health)	10
Injuries and poisoning	11
Influenza	20
Asthma in under 45s	30
Stroke	40
Heart disease	50
COPD	60
Diabetes	70
Teenage pregnancy	80
Mental health	90
Affective disorders	91
Schizophrenia	92
Alcohol and drug abuse	93
Other mental health disorders	94

Figure 49: Table of diagnoses groups

Once again the method outlined in Chapter 5, section 5.2 was used to calculate risk admission scores for each neighbourhood Type in England. The same method employed in section 7.2.2 was then used to develop the practice profiles. For each practice in Camden, the percentage of registered patients per Mosaic Type was calculated. These percentages were then multiplied by the previously calculated index value for each neighbourhood Type. The final step involved summing the results for each practice to give an overall index value, highlighting the level of likelihood for each health outcome for each practice, according to their registered resident population.

7.3.3 Results

For each practice the likelihood of risk to particular diseases was calculated, accounting for the demographic profile of the patient register. The results indicated the propensity for patients registered at particular practices to be admitted to hospital for certain disease and thus indicating risk likelihood in the population. The analysis highlighted certain diseases which showed a wider distribution of index values, as calculated by their standard deviation.

For developing communication and promotional campaigns large standard deviations of index values are preferable because they are indicative of differences existing across different demographic Types. Both efficiency and effectiveness of promotional and social marketing campaigns can be maximised by focusing on those practices with high index values for diagnoses and where the standard deviation for the diagnoses across all practices is large.

Rank	Diagnosis	Standard Deviation
1	Schizophrenia	57.00
2	Drug & Alcohol Abuse	34.85
3	Teenage Pregnancies	33.85
4	Asthma	33.31
5	Mental Health	32.01
6	COPD	23.67
7	Influenza	21.71
8	Affective Disorders	21.50
9	Other Mental Health	20.40
10	Lung Cancer	16.83
11	Diabetes (all)	16.50
12	Cervical Cancer	16.29
13	Heart Disease	10.31
14	Emergency	8.82
15	Bowel Cancer	8.63
16	Breast Cancer	8.42
17	Stroke	8.09
19	Cancer	6.18
20	Injuries & Poisons	6.06

Table 25: Standard deviations of practice index values for different diagnoses

Table 25 lists the standard deviations of the composite risk index values for practices in Camden. The diagnoses highlighted in red are practices with the top 10 largest standard deviations. The widest distribution across the practices pertains to schizophrenia related illnesses. This indicates patient tendency to be admitted to hospital for schizophrenia is not evenly distributed across Camden practices. The width of the distribution is more than nine times larger than that of injuries and poisons. Drugs and alcohol abuse, teenage pregnancies, asthma and mental health all have standard deviations greater than 30 and thus appear suitable for wider social marketing campaigns. Each of these conditions is related to wider 'social problems' and in the medical literature are commonly linked to low income, poor housing conditions, educational attainment and material deprivation.

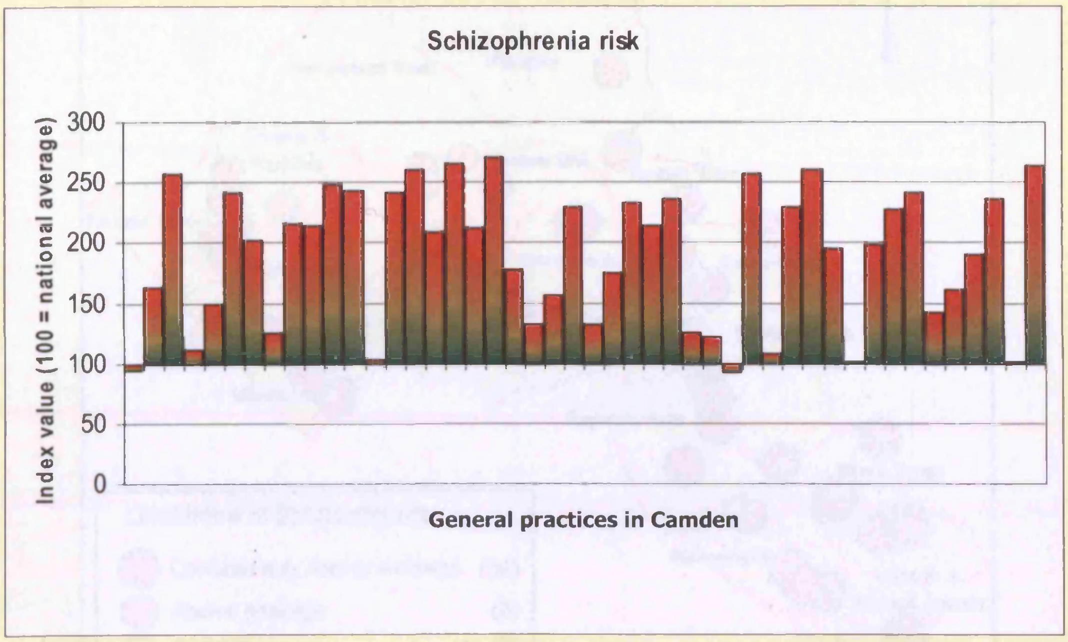


Figure 50: Schizophrenia risk for patients registered with general practices in Camden

Figure 50 highlights the distribution of schizophrenia risk for patients registered at general practices in Camden. The graph highlights the deviation of risk from the national average of 100. All practices with composite index scores for schizophrenia above the national average of 100, indicate patients who have an elevated risk of being admitted to hospital for schizophrenia.

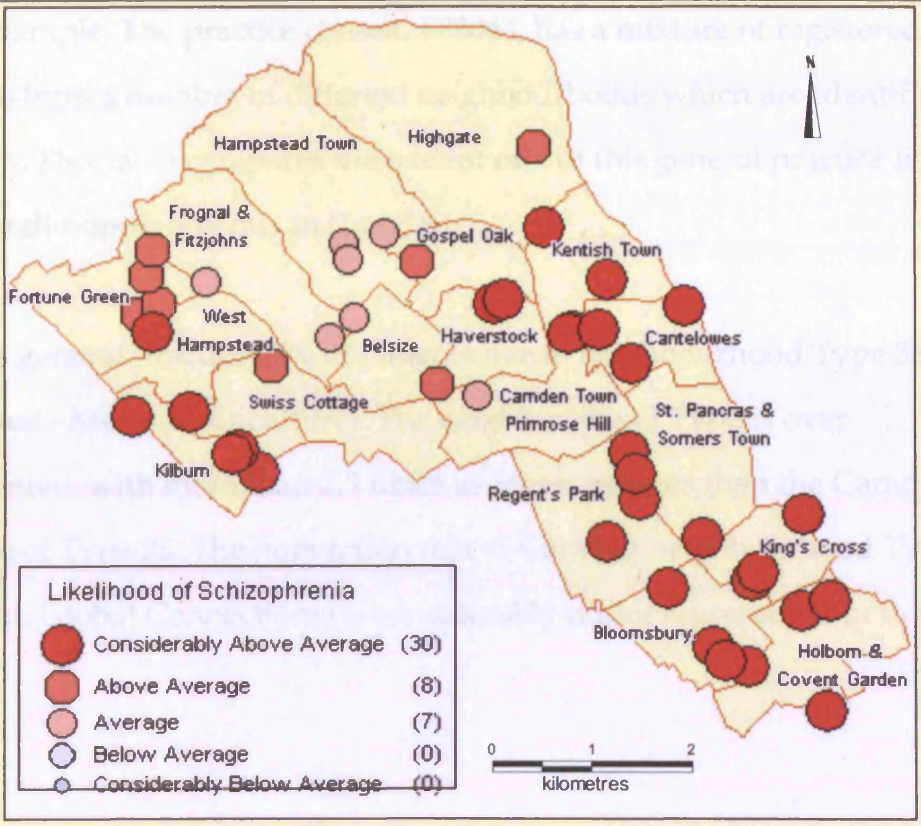


Figure 51: Map of Schizophrenia risk patients registered at general practices in Camden

The distributions of the risks index scores were then mapped to identify their spatial distribution. The output of this can be seen in Figure 51. The map provides a spatial representation of the results outlined in Figure 50. A graduated symbol map with a colour gradient has been used to highlight the risk scores. For practices with scores considerably above the national average, a large dark red shade has been used. The same categorisation scale as that used in section 7.2.1 was applied. Practices with the wards of Frognal and Fitzjohn and Hampstead Town are the only ones to have likelihood in line with the national average; all others are above average. Is this because of the high rates of drug abuse which is frequently associated to drug taking or is it linked to the developed mental health services that exist in Camden? This certainly points to a requirement for further work.

Individual practice profiles can be created using the proportion of patients registered at each practice. A practice in the south of Camden has been used as an example. The practice chosen, F83044, has a mixture of registered patients from a number of different neighbourhoods which are identified in Table 26. This table compares the patient mix of this general practice list to the overall population mix in Camden.

For this general practice 55% of patients live in neighbourhood Type 36 (deprived - Metro Multiculture). The neighbourhood Type is over-represented, with more than 2.5 times as many patients than the Camden average of Type 36. The population mix of Camden neighbourhood Type 01 (affluent, Global Connections) is considerably under represented in Practice F83044.

Type	Percentage of Patients in Practice: F83044	Percentage Camden Population	Index neighbourhood type for Practice F83004 versus Camden	Percentage of Patients in Practice: F83055	Index neighbourhood Type for Practice F83055 versus Camden
01	0.02	0.22	8	0.27	121
02	0.00	0.05	0	0.08	175
03	0.00	0.00	0	0.00	0
08	0.01	0.01	87	0.00	57
26	0.01	0.00	682	0.00	0
27	0.00	0.01	0	0.01	111
28	0.18	0.25	69	0.14	54
29	0.07	0.12	53	0.32	256
30	0.00	0.02	0	0.14	680
33	0.05	0.02	267	0.01	41
34	0.02	0.02	149	0.00	0
35	0.03	0.01	443	0.00	0
36	0.55	0.25	218	0.02	8
38	0.02	0.01	264	0.00	6
39	0.03	0.01	500	0.00	73
40	0.00	0.00	0	0.00	0
48	0.00	0.00	0	0.00	0
49	0.00	0.00	0	0.00	0
50	0.00	0.00	0	0.00	0
51	0.00	0.00	0	0.00	0
52	0.00	0.00	0	0.00	0

Table 26: Geodemographic profile for Camden and Practice F83044

The demographics of the patients registered with this practice should in essence determine the case-mix of diagnoses. Using the index scores, practice profiles were developed. A radar diagram for each general practice was used to represent composite index scores for each of the diseases investigated and listed in Table 25. The radar diagram structures the conceptual space of health inequalities along vectors representing variables which correspond to indicator scores of different predicted health outcomes. They are a useful visual tool that facilitates relative comparisons of health outcomes. For each GP practice the radar diagram displays the current predicted level of patient need for a range of conditions. Such an example is illustrated in Figure 52 for the Bloomsbury Surgery, in the locality of Hunter Street in south Camden (Practice code: F83044). The conceptual space representing expected inequalities in health for the practice is denoted by plotting the expected health risks for 20 different diseases and conditions. The national average for each condition is set to 100, and is identified on the diagram by the orange circle. Risk scores located inside the orange circle are below the national average, and intervention planning and social marketing efforts should be concentrated on the outcome variables that are plotted outside the orange circle. Figure 52 indicates that because of the patient mix, the practice is expected to have above average episodes of drugs and alcohol abuse, schizophrenia, teenage pregnancies, mental health issues, coronary obstructive pulmonary disease (COPD), influenza, cervical cancer, asthma and affective disorders (mental illness conditions that affect moods such as bi-polar).

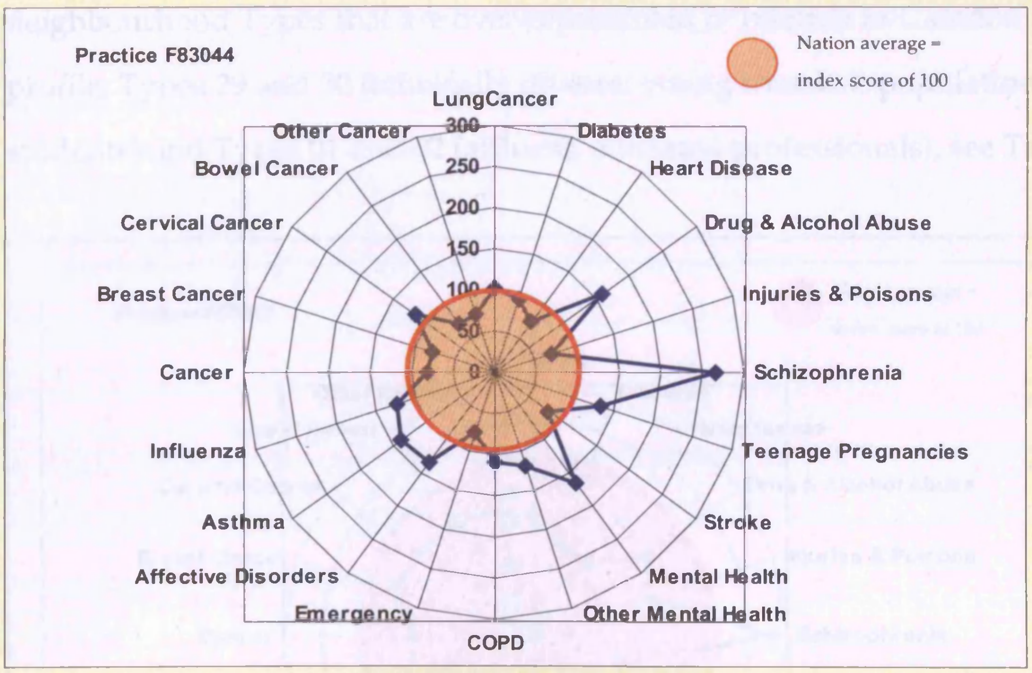


Figure 52: Practice profile for practice code F83044

The standard deviation was calculated for the distribution of risk scores for each of the outcome variables for this general practice: the result was 46.9. This value indicates a relatively wide distribution of risk scores and consideration should be made for those health outcomes and conditions at the extremities of the distribution with values above the national average which are "outliers". They have higher than expected scores and deviate excessively from national expectations. Thus, for this practice, social marketing campaigns and interventions are appropriate, particularly for the following conditions; asthma, influenza, breast cancer, mental health, teenage pregnancy together with drugs and alcohol. The social determinants of these conditions and illnesses can be caused by health- impacting behaviours and are therefore suitable candidates for social marketing.

To demonstrate the influence of patient-mix on the types of priorities that practices face, a general practice with a contrasting geodemographic profile to the previous example was chosen. The comparative practice, F83055, is located in the centre of the borough. This practice has a number of neighbourhood Types that are over-represented in relation to Camden's profile; Types 29 and 30 (ethnically diverse, young transient populations and students) and Types 01 and 02 (affluent, educated professionals), see Table 26.

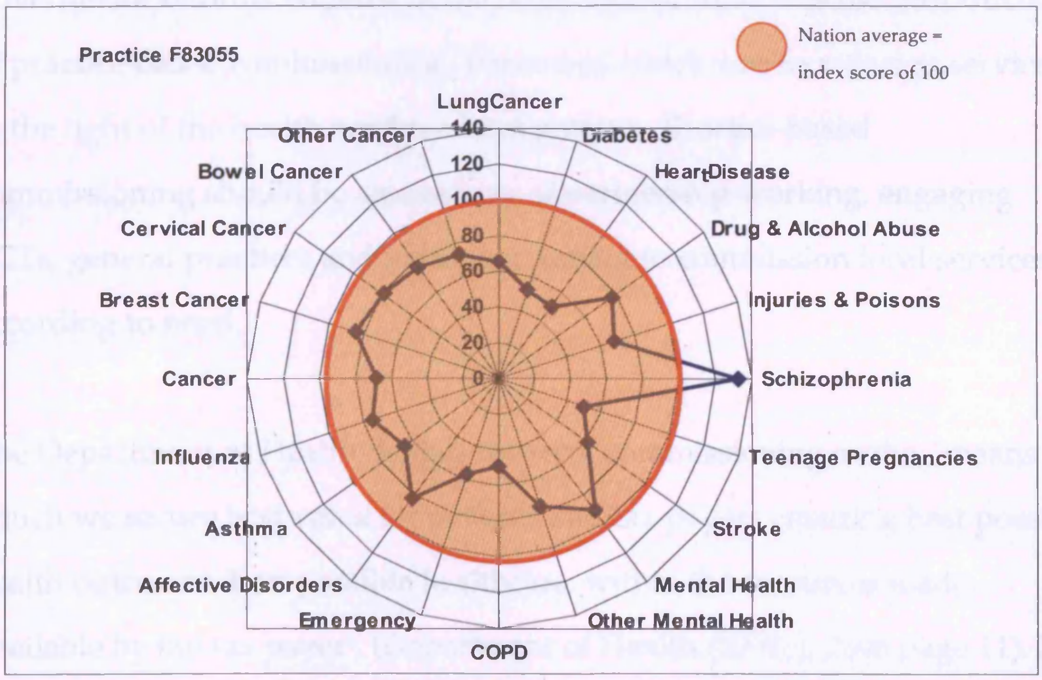


Figure 53: Practice profile for practice code F83055

Close inspection of Figure 53 for this general practice, indicates that all but one of the health outcome risk variables have scores well below the national average (once again marked by the orange circle). The only condition with higher risk than expected is schizophrenia. Once again, the standard deviation was calculated for the distribution of risk scores for this practice, and the result was 18.9, representing a very narrow distribution of scores. For this practice, the patient-mix corresponds to an overall healthier patient list as compared to the other practice. This practice is much less likely to have patients admitted to hospital for the range of conditions and illnesses analysed. Practice based social marketing initiatives for this practice may not, therefore, be a priority or an efficient use of resources as it is unlikely to return maximum results for the amount of effort.

7.3.4 Discussion

Since the Labour government came to power in 1997 there has been a raft of government reforms targeted at the NHS. One of these was the introduction of practice-based commissioning, the aim of which was to redesign services in the light of the health needs of local patients. Practice-based commissioning should be an example of partnership working, engaging PCTs, general practices and local government to commission local services according to need.

The Department of Health defines the term commissioning as the, 'means by which we secure best value for patients and tax payers ensuring best possible health outcomes, best possible healthcare, within the resources made available by the tax payer', (Department of Health (2004c), 2006 page 11). In line with these commissioning goals the Department of Health also released a white paper in 2006 corresponding to the importance of community based services; 'Our health, our care, our say' (Department of Health (2004a), 2006, chapter 2), stressing the requirement to commission for health and well-being to ensure health improvement is at the heart of the commissioning process.

There is significant importance surrounding the identification of local health care burdens for different communities, in line with commissioning of services. For services to be commissioned effectively and interventions planned appropriately, the patient mix of general practices must be understood. In the case of Camden, the polarised nature of different communities does not occur at across large geographical extents. The contrasts in the location of population sub-groups manifests in a geographically localised manner, where a matter of 10s of metres can result in a stark change of population and by corollary a dramatic shift in health outcomes. The differences in population sub-groups indicated by the geodemographic classification were highlighted at the beginning of this chapter and are returned to in Chapter 8. At the beginning of this chapter the map of neighbourhood Types in Camden illustrated the spatial proximity (indeed neighbouring) of different neighbourhood Types (conducting a nearest neighbour clustering analysis would evaluate this quantitatively). The health care burden for practices differs according to where they are located in the borough, how they define their catchment areas and location of the population characteristics of surrounding local residential neighbourhoods.

This means that general practices have diverse populations registered on their list and through understanding the location of different population sub-groups and the geodemographic profiles, of patient registers, PCTs are better places to commission services and conduct effective campaigns. The examples illustrated in this case study highlighted the diverse population profiles of just two practices in Camden. The combination of geodemographic analysis with hospital episode data to create local profiles of risk, representing patient-mix, for different conditions and illnesses provided a useful method for exploring the conceptual space of health inequalities and status at the level of the general practice. Furthermore, if a clustering algorithm was applied to the general practices they could be grouped together according to their predicted health outcome scores and their geographical proximity. This would enable shared services and campaigns to be effectively planned and managed across similar practices.

Further exploratory work would be useful to determine the usefulness of these techniques as an alternative framework for creating estimates of actual disease prevalence in population sub-groups. Comparisons with data derived from the Quality Outcomes Framework (QOF) could be used to compare estimations by geodemographic Type and prevalence rates for practices, but would be restricted to the number of conditions recorded in the database and the quality the dataset.

The results throughout this section highlighted that magnitude of variation in all diabetes and more specifically diabetes Type II across different neighbourhood Types and the importance of these differences for informing social marketing campaigns and interventions. This research is complemented by research conducted by Cox et al. (2007) who noted that Type 2 diabetes linked to deprivation, and was more common in deprived neighbourhoods but less common in deprived neighbourhoods surrounded by less deprived areas. The research was conducted at the scale of the output area. Further analysis using the neighbourhood risk techniques to identify socially segmented neighbourhoods together with a LISA analysis of the how similar diabetes risk is compared to its neighbourhood would complement the research in both papers. The micro-scale patterns of diabetes for functional neighbourhoods and the impact of the surrounding social space could then be evaluated.

7.4 Exploring neighbourhood as a predictor of response to breast screening services

The case studies discussed in this chapter and the previous chapter have presented the geodemographic framework as a viable alternative to explore, describe and understand differences in health outcomes at the scale of the local neighbourhood. The previous chapter combined national survey data with a national geodemographic classification to view compositional and contextual differences in lifestyle related health-harming behaviours. The first two case studies in this chapter substantiated the framework by ascribing geodemographic characteristics to national datasets of hospital admissions to explore diabetes risk. This enabled the scalability and validity of the indicators to be evaluated and quantified. The final section of this chapter moves away from predicting disease or illness risk to understand the framework in the context of operational health data and service uptake from the perspective of cancer screening. It is conducted by adopting the methods developed throughout Chapter 6 and the first part of this chapter and in essence consolidates the learning outcomes.

7.4.1 Introduction

The final case study in this chapter combines geodemographic techniques with the measurement of service outcomes to enrich operational data. It investigates the feasibility of providing health professionals with a greater insight about service uptake through the provision of valid measurements on which to base service planning, delivery and promotional campaigns. The aim of the case study was to provide an alternative framework for measuring service outputs, building upon the knowledge gained through the previously presented case studies which could then be used by health professionals to improve uptake of screening services in Camden. In this example geodemographic classifications provide proxy information for exploring the neighbourhood context.

As already set out in the previous case studies, geodemographics do not set out to explain the associations between health and place, but they do enable the incorporation of group level variables into analysis. They provide a multivariate classification based on small areas, equivalent to neighbourhoods. Analysis incorporates the effects of these variables defined at a higher group level, on outcomes that are defined at the lower individual level. This research has described an alternative measurement system for outputs that incorporates both individual and group level variables and in this case study, applies it to transactional operational health data.

The purpose of this case study was to demonstrate the use of combining geodemographic and geographic analysis to divide the population into different Types and Groups providing contextual information about the neighbourhood. The dissection of the data in this manner should further the understanding of the socio-economic and behaviour characteristics of women who have not responded to calls for breast screening. The objective was to identify population sub-groups for targeting through social marketing initiatives to improve screening uptake.

7.4.2 Problem definition

Cancer screening programmes are required to ensure early disease detection. Detection is considered to be the most effective way to reduce morbidity and mortality from breast cancer (Broeders and Veerbeck, 1994). Structured screening programmes have been proven to be more effective in reducing the rate of death from breast cancer than sporadic screening of selected groups of women (International Agency for Research on Cancer, 2002). Thus the implementation of organised, regular call and recall screening for all members of an age cohort is a more appropriate way to reduce the impact of breast cancer rather than targeted initiatives at just specific population groups, as it is more effective in enabling early detection of cancer cells.

Since 1987 free breast screening has been available in the UK to all women between the ages of 50 to 64. Since the implementation of the programme the age cohort of eligible women has increased. Currently all women between the ages of 50 and 70 are invited to attend screening. To ensure the success of the NHS Breast Screening Programme (NHSBSP), the Department of Health and the NHSBSP set a national target, a performance measure that all Primary Care Trusts (PCTs) should achieve. The target states that a coverage rate exceeding 70% must be achieved by each primary care trust to ensure the efficacy of programmes. The coverage rate is a quantitative measure that defines the proportion of eligible women who have received screening at least once in the previous three years (the period of a screening round). It is different to the measure of uptake, which measures the proportion of invitees who attend screening.

If the proportion of eligible women attending screening exceeds this target value of 70%, then early presentation of breast cancer combined with effective treatment will be achieved, and consequently disease morbidity and mortality rates should reduce with time, and the programme will be deemed effective. By the end of March 2004, 74.9% of the women aged 53 to 64 who were registered with general practices in England had been screened at least once in the previous 3 years. Due to the 3 year cyclic nature of the screening invitations, coverage rates are calculated for women aged 53 to 64 on the 31st March 2004.

A total of 247 of the 303 primary care organisations in the UK achieved a coverage of 70% or higher. Camden, the primary care trust (PCT) of focus in this research, is one of the 24 primary care organisations whose breast screening coverage was lower than 60%, a pattern repeated across many of the London PCTs (NHSBSP, 2005).

The coverage rate for Camden PCT at the end of March 2004 was 51%, as it was in 2003. Figure 54 shows the coverage rates for women aged between 53 and 64 for the PCTs that comprise the North Central London Strategic Health Authority (NCLSHA). The data highlight that the Inner London boroughs of Camden and Islington experienced the lowest coverage rates over the 3 year period. The coverage rates for Camden are the lowest within the NCLSHA, with its neighbouring borough of Islington also having a coverage rate below 60% at the end of March 2004.

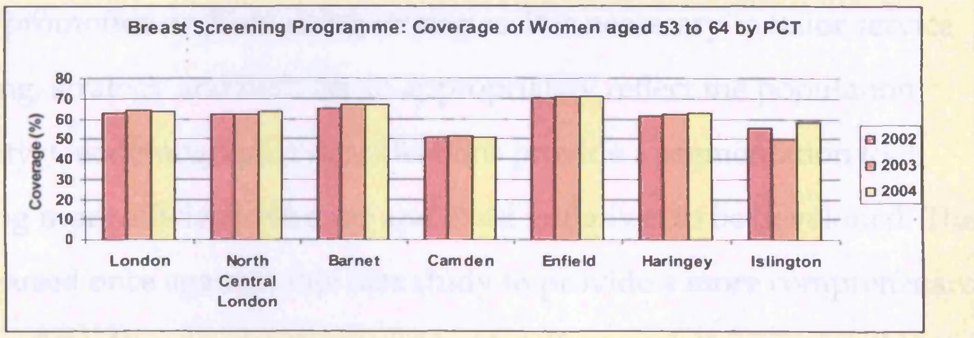


Figure 54: Breast Screening Programme: coverage of women aged 53 to 64 for North Central London PCTs

In 2004, 5,523 women in Camden were screened during the three year coverage period. To reach the national target of 70%, 7,572 should have undergone breast screening, 2049 more women than were actually screened.

This low coverage rate led to a joint initiative between Central and East London Breast Screening Services (CELBSS), Camden PCT and General Practices to improve the screening rates. The Camden recall project was set up to recall all women who did not respond to calls for breast screening between January 2003 and September 2004.

Only a limited level of information is stored in the non-responder's dataset, merely the NHS number, date of birth (from which the age can be extracted), address and full unit postcode of each woman are recorded. No data are available about the socio-economic conditions or lifestyles of the women, and thus service planning and delivery according to need can be difficult to predict and achieve. Comprehensive, accurate and robust data enable service evaluation improvement, as well as planning. Data-driven information analysis facilitates knowledge gathering and provides an evidence base for strategic policy and planning.

To ensure that services are efficient and serving the population in need, it is necessary to understand that different neighbourhoods may require different health promotion and screening strategies. It is necessary to tailor service planning, strategy and delivery to appropriately reflect the population diversity. Geodemographic classifications provide a segmentation tool, enabling more efficient, effective and exact initiatives to be developed. They will be used once again in this case study to provide a more comprehensive picture of the Camden residents and to identify population groups that are less likely to access screening service. Through the use of geodemographic analysis this case study demonstrates the plausibility of enriching locally available operational datasets to add contextual understanding of population differences between different neighbourhoods.

The application of geodemographic analysis to screening data should lead to the description and understanding of pattern variations of poor uptake at a neighbourhood level and provide added-value information to those engaged in service delivery and better planning. More detailed comprehension of population diversity will assist PCTs in the creation of opportunities through provision of support and information to enable people to make informed health decisions and improve service compliance.

7.4.3 Data, method and analysis

Three key datasets were used to describe the variation among neighbourhood groups of women who do not respond to invitations for breast screening. First, the list of non-respondents for the time frame of the study, a list of all eligible women who did not attend screening between January 2003 and September 2004 and had not responded to any of the calls to attend. The second dataset used was a list of resident registered patients for each Camden practice. Finally, the third dataset used was the geodemographic classification for each postcode within the Trust's boundary. The investigation was conducted for women not attending breast screening services in Camden. Mosaic (Experian, Nottingham) was used to categorise women who had not presented for breast screening, following routine invitations requesting their attendance. Each woman who had not attended the service was assigned a neighbourhood Type and Group, according to the postcode unit of their place of residence, the unit of analysis for this investigation.

A list of all non-respondents was extracted from the Exeter database (an NHS-wide database holding a variety of patient data) and was sent to general practitioners to update; women who were known to have moved out of the Borough, women who had since died and women for whom screening was no longer appropriate (for example women whom have had a double mastectomy) were removed from the list.

The final updated non-respondents list was used to recall these women for screening. The invitation procedure involved sending an open appointment letter to each woman. Each was requested to call a telephone administrator who arranged an appointment at the most appropriate time for the woman. Under this new initiative, Camden PCT had a designated administrator to talk personally to the women when they called up to make their appointment (in the past there was an automated answer phone), and the invitation letters were personally signed by their general practitioner (as opposed to a random senior manager within the PCT). It is upon this dataset that the analysis was conducted.

The analysis was carried out in two stages. Firstly, basic spatial analysis was conducted using administrative boundaries that are commonly used during service delivery and planning. The second stage involved exploratory data analysis using neighbourhood analysis (geodemographic) techniques to understand differences in the service uptake. The first stage of the non-respondent analysis used the postcode as the geographic identifier to link women to administrative boundaries. The postcode units for the residential addresses of the women were linked to a unit postcode dataset for all of Camden. The use of the postcode enabled the datasets to be spatially referenced, and the address of the non respondents to be mapped. A density map was created to highlight locations within the PCT boundaries that had a higher density of women who did not respond to calls for breast screening (Figure 55).

The second phase of the analysis created neighbourhood (geodemographic) profiles for the different Groups of people living within the postcode units of

Camden. Through the use of the postcode it was possible to link the Groups and Types to the women who have not responded to calls for breast screening. This enabled a more discriminatory analysis to be completed, allowing investigation by different socio-economic Types and lifestyle.

As explored in the previous case studies, neighbourhood (geodemographic) analysis uses the technique of calculating index values for different behaviours for each neighbourhood Group or Type as a means of measuring likely outcomes or expected behaviour. In the previous case studies their relevance to predicting health outcomes was identified. The same method as all the previous case studies was used to create index scores of neighbourhood uptake to screening services. The process was outlined in Chapter 5, section 5.2.

Unlike the previous case studies indictors were created for the hierarchical neighbourhood Group and not the neighbourhood Type, because the cohort under investigation was quite small and when cross- tabulated with neighbourhood Types would have resulted in some very small counts which would render the indicators invalid (they are very sensitive to small numbers). The postcode of each non-respondent was extracted and used as a unique identifier and the geodemographic classification ascribed accordingly. The number of non-responders for each geodemographic Group was totalled and the index value created. The resultant index value represented the likelihood of attendance at screening services.

This permitted the comparison of service utilisation and equity of access for different groups of residents across Camden. These indices were then applied to general practice registration lists following the method presented in section 7.1.3. This enabled the calculation of index values representing the likelihood of screening uptake for general practices.

7.4.3.1 *Results part 1; density map of non-responders*

Initially, a density map was created, showing the areas where women were likely to attend screening. This type of map has been used as it maintains anonymity; individual point locations are not available on the output map. The 'hot spots' were based upon a point pattern that represents the density of non-respondents as they vary across Camden. Figure 55 highlights areas where the concentration of women not responding to calls for breast screening was greater, relative to the distribution of non-respondents across the whole of Camden. If the hotspot is dark red it signifies the greater the number of women in that area who do not present for breast screening. In reality this map is an artefact of population density, as it does not indicate true hotspots as it has not been standardised.

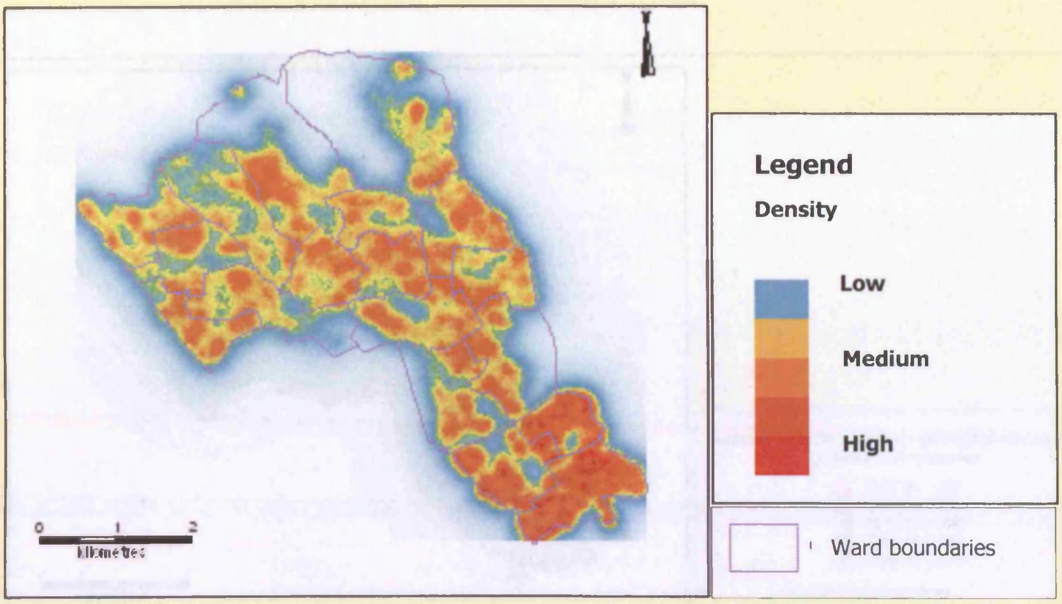


Figure 55: Density map showing residential hotspots of non-responders to breast screening

To illustrate the differences between areas where residents do not respond to screening, invitation acceptance rates were calculated for small areas. The heterogeneity of the spatial patterning of the residential locations of women not responding to calls for breast screening is further illustrated by Figure 56, which shows the breast screening non-respondent rate per 1000 women for output areas across Camden.

The measurement rate per 1000 women was calculated because government crime and health statistics are commonly reported in this fashion. Once again it is not useful because output areas are relatively heterogeneous and do not provide sufficient detail about service uptake to enable specific targeted interventions to be developed.

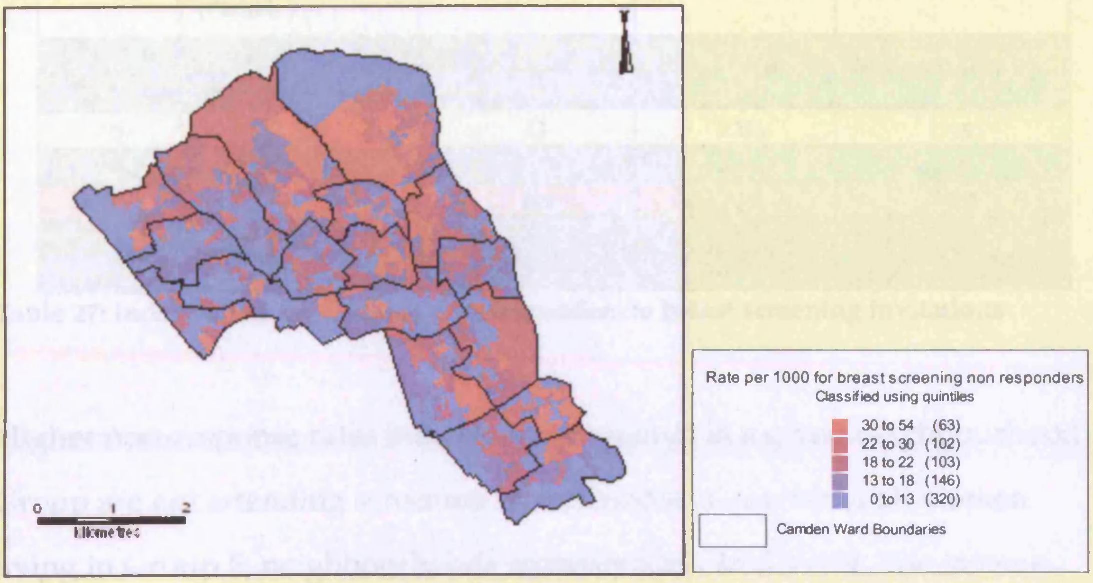


Figure 56: Rates per 1000 women aged 50 to 65 who have not responded to calls for breast screening

7.4.3.2 *Results – part 2; neighbourhood profiles:*

The results of the non-respondent index values for Camden were analysed. Table 27 lists the results for each neighbourhood Group together with the 'non-response rate'. This measure represents the proportion of patients per neighbourhood Type that had not attended breast screening in the given reporting period. The non-response rate is the ratio of non-respondents compared to the number of female registered patients (eligible patients) in the age cohort. It is a useful measure of success, as it shows how the service is performing across different neighbourhood Groups and it indicates how equitable the service is.

Group	Number of Camden registered patients aged 50 to 65 (Female)	Number of Non-Respondents in Camden	Non- response rate	Non-response Index
A	3787	880	23%	96
B	15	2	13%	55
D	100	23	23%	95
E	3987	961	24%	99
F	3075	784	25%	105
I	60	17	28%	117
J	45	10	22%	91

Table 27: Index values for Camden non-responders to breast screening invitations

Higher non-response rates indicate more women in a given neighbourhood Group are not attending screening. For example, a quarter of all women living in Group F, neighbourhoods representative of diverse, low income, public housing, do not attend screening services. Indeed they are 5 % more likely not to attend screening services than the Camden average. As a large proportion of Camden's cohort is in this neighbourhood Group, this suggests that women in this Group would benefit from social marketing campaigns and further observation to understand their barriers to screening. Furthermore, women in this neighbourhood Group are more likely to suffer from higher morbidity (illness) because women in low-income population groups present with cancer at later stages where the illness is more advanced. Early detection through screening would result in lowering morbidity for this population cohort.

To determine how equitable the service is once again we look at the non-response rates. In this example all women in the age cohort should be attending screening regardless of their socio-economic characteristics and vulnerability. If the non-response rates for each neighbourhood Type show considerable variation, then the equitability of the service is low. For the service to be considered equitable the rates should be consistent across the neighbourhood Groups. Figure 57 shows that Camden breast screening service is equitable in the groups that it reaches. Although, this is not an absolute an alternative possibility for the appearance of equity is at that at this level of aggregation (population Group), the neighbourhood classification may not be a suitable discriminatory tool, because of the small numbers and overall low-response rate in Camden.

Care must be taken when interpreting the non-response rates. It was previously mentioned that the national target for screening coverage is 70%, and this figure is soon to rise to 80%. If the non-response rates in Table 27 were reversed, implications could be made concerning the overall performance of the breast screening service. The rate for Group A (Symbols of Success) is calculated at 23%, suggesting that 23% of women living in this neighbourhood Group do not respond to invites for breast screening. The reverse of which implies that 77% of people living in this neighbourhood Group actually do attend screening. These statistics cannot be compared to the national target for a number of reasons. Firstly, this case study is exploring poor uptake of service use, not coverage rates. Accessibility issues relating to the database which stores the screening uptake information meant that the data used to create the index dominator (ie the baseline) is only a proxy for the number of eligible women. The actual baseline data listing the total number of women who were invited for screening was not made available to this research.

The response rate is the index value representing the likelihood of a woman in a particular neighbourhood Group to invitations for breast screening. Figure 57 shows the index values for different demographic Groups in Camden. It shows that compared to the Camden average (index value=100), the neighbourhood Groups of women that are most likely not to respond to a call for breast screening. Two Groups in particular show higher than average rates of non response: women living in postcodes represented by Group I

(Twilight Subsistence) who are typically elderly people subsisting on meagre incomes and living in council accommodation (perhaps they have difficulty physically getting to the service); and women in Group F (Welfare Borderline), people living in this postcode are mostly reliant on the council for accommodation and benefits.

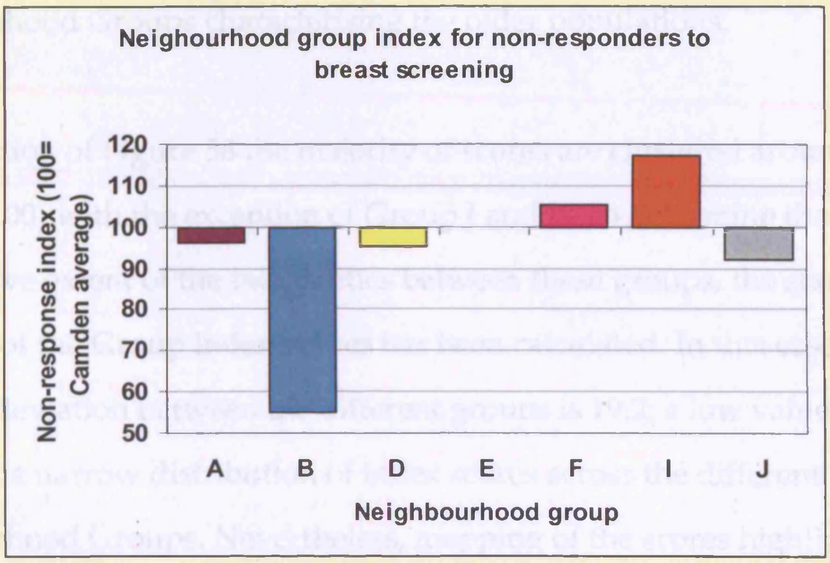


Figure 57: Neighbourhood Group index for non-responders to breast screening invitations

Group A (Symbols of Success), shown in purple on the chart, has an index value of 96. Group E (Urban Intelligence), has a calculated index value of 99 and finally the index value for Group F (Welfare Borderline) is 105. Each of these Groups has different socio-economic and lifestyle characteristics. Group A is representative of people with successful careers who live in sought-after locations. Women living in Group A are most likely to be affluent. The neighbourhood Group E is most likely to be comprised of well educated people of average affluence. Women in Groups I are 20% more likely not to respond to screening invites than the average neighbourhood response for Camden.

At this point it is important to note that the results appear to be skewed; more index scores are below 100 than above. This is the result of two factors. The first is that despite using neighbourhood Groups as opposed to Types, the indicator scores are still subject to the restrictions of small numbers in the denominator (see Group B). The second is that the population pyramid for Camden is skewed towards a younger population below 45 years of age. One would expect those neighbourhood Groups (like B and E) to have few women in the relevant screening cohort, whilst overall in Camden there are fewer neighbourhood Groups characterising the older populations.

On inspection of Figure 58 the majority of scores are clustered around the average (100) with the exception of Group I and F. To determine the comparative extent of the inequalities between these groups, the standard deviation of the Group index values has been calculated. In this case the standard deviation between the different groups is 19.2; a low value indicating a narrow distribution of index scores across the different neighbourhood Groups. Nevertheless, mapping of the scores highlights the spatial extent of neighbourhoods that are not likely to respond to invites to breast screening.

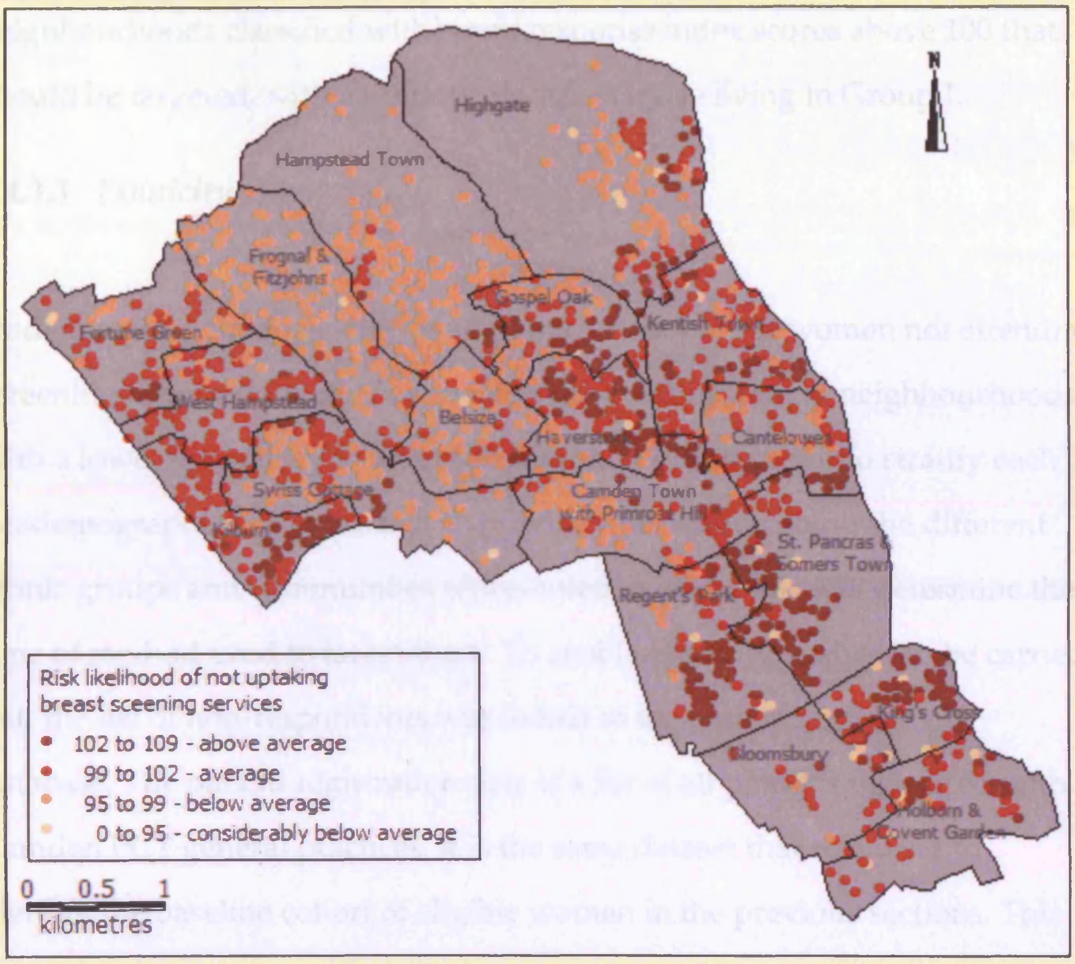


Figure 58: Map of neighbourhood Group index values for Camden representing poor likelihood to uptake to breast screening services

Figure 58 illustrates the spatial distribution of the index values across the Borough, while the ward boundaries are used as the background layer. The data have been classified into quartiles. It is clear that above average rates of non-respondents were to be expected in both the centre and the south of the Borough and by varying degrees to the North West. Differences in breast screening uptake do occur over space, although this type of analysis does not confirm whether these spatial patterns are statistically significant. Once again the clustering of high values could be further enhanced by an analysis of spatial autocorrelation, to indicate where neighbourhoods with high likelihood of non-response are located near other neighbourhoods with high likelihood of non-response. For the purpose of social marketing it is the neighbourhoods classified with a non-response index scores above 100 that should be targeted, with emphasis placed on those living in Group I.

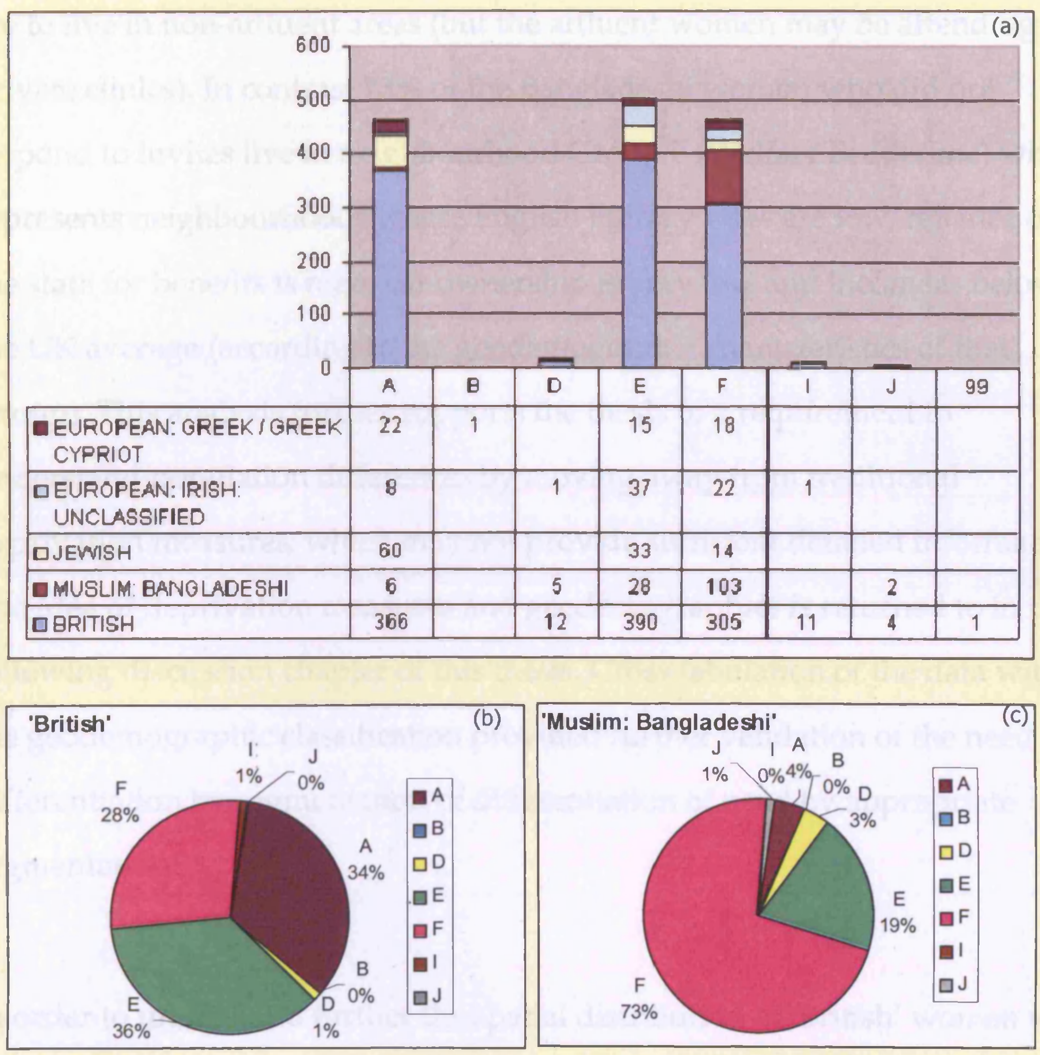
7.4.3.3 Ethnicity

Understanding the socio-economic characteristics of the women not attending screening is important and has enabled the identification of neighbourhoods with a lower propensity to attend screening. It is also useful to stratify each geodemographic Group further to provide information about the different ethnic groups and communities represented, because this will determine the type of method used to target them. To enable ethnicity analysis to be carried out, the list of non-respondents was linked to the patient registration database. The patient registration data is a list of all patients registered with Camden PCT general practices. It is the same dataset that was used to identify the baseline cohort of eligible women in the previous sections. This dataset contains information relating to surname, forename, date of birth, gender, address and place of birth. It has been used to derive a proxy classification for ethnicity.

Comprehensive details and discussion about this proxy ethnicity classification is beyond the scope of this paper and is discussed in depth by Mateos (2007, 2006), but a brief overview is required. The place of birth field within the patient registration dataset is a free text field. These data are often inconsistent and may be of poor quality: for example, data in the place of birth field ranged from "at home" to any of 31 different spellings of Bangladesh. The field, where completed, was standardised using a variety of rules to enable the country of birth to be ascertained. The next step involved first classifying the surname and then the forename according to their linguistic provenance. This information was then combined to create a model known as CEL, which is a proxy indicator for ethnicity according to a classification of culture, ethnicity and language (CEL), (Mateos et al., 2007).

Through matching the NHS number (which is unique for every patient) of the non-respondents to the patient register, ethnicity analysis was carried out using the CEL classification. A total of 2186 non-respondents to breast screening invitations could be linked to the patient registration database using unique NHS identifiers. Thus, 81 % of the non-respondents dataset could be assigned an ethnicity classification code.

Figure 59 (a) lists the five ethnic groups that contained the most women who did not respond to breast screening invitations. 67% of the non-respondents (who could be assigned a CEL code) were accounted for within the top 5 most common ethnic groups representative of Camden's population; British, Muslim Bangladesh, Jewish, Irish and European Greek-Cypriot. Furthermore within the most frequently occurring ethnicities 74% of the non-responders were classified as "British".



	A	B	D	E	F	I	J	99
EUROPEAN: GREEK / GREEK CYPRIOT	22	1		15	18			
EUROPEAN: IRISH: UNCLASSIFIED	8		1	37	22	1		
JEWISH	60			33	14			
MUSLIM: BANGLADESHI	6		5	28	103		2	
BRITISH	366		12	390	305	11	4	1

Figure 59: (a) Table of top 5 ethnic groups who did not respond to breast screening invites and pie charts of percentages of non-respondents per geodemographic Type for British (b) and Bangladeshi (c)

The ethnic group labelled 'British' was the largest group of non -responders. This information was cross-referenced according to the neighbourhood (geodemographic) Group in which each woman resided, Figure 59 (a). Within the group of British women no one particular neighbourhood Group had predominantly greater proportions of non-responders; 34% of British women in Groups A, 36% in Group E and 28% in Group F did not attend screening as highlighted by Figure 59 (b). British women who do not respond to calls for breast screening are just as likely to live in affluent areas (Group A) as they are to live in non-affluent areas (but the affluent women may be attending private clinics). In contrast 73% of the Bangladeshi women who did not respond to invites live in neighbourhood Group F (Welfare Borderline) which represents neighbourhoods where English literacy rates are low, reliance on the state for benefits is high, car ownership is very low and income is below the UK average (according to the geodemographic characteristics of that Group). This analysis further supports the thesis of a requirement to understand population differences by moving away from traditional deprivation measures, which may not provide sufficient detailed information. The idea of deprivation measures and geodemographics is returned to in the following discussion chapter of this thesis. Cross tabulation of the data with the geodemographic classification provided further validation of the need for differentiation by segmentation of differentiation of need by appropriate segmentation.

In order to understand further the spatial distribution of 'British' women who have not responded to breast screening invites, the density maps shown in Figure 60 were created. The maps again show spatial differences between the locations of residence of the different neighbourhood Groups. The 'hotspots' for non-responders that are classed as neighbourhood Group A (Figure 60 a) (Symbols of success) are found to the northern central part of the Borough, within and around Hampstead and Parliament Hill (on the borders of Hampstead Heath) and a little further to the south near the borders of Primrose Hill.

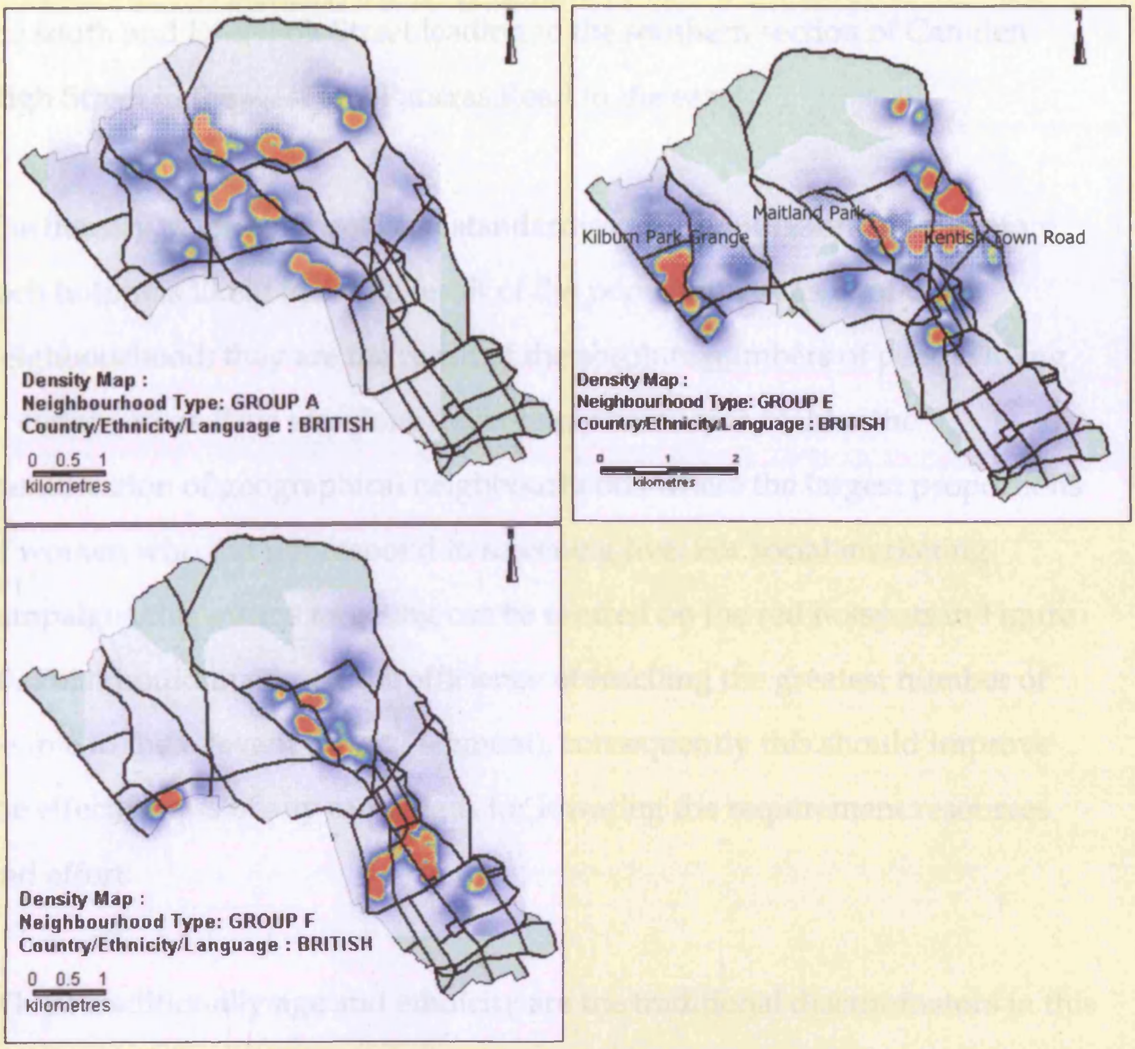


Figure 60: Location of non-responders with a Cultural/Ethnicity/Language classification of 'British'

The 'hotspots' of British non-respondents living in group E (Figure 60, b) are found in the east and west of the Borough. In the west the hotspot is found tangential to Kilburn Grange Park. In the east of the Borough there are 3 hotspots. The first is within Maitland Park following Maiden road, the second and third are found on both sides of Kentish Town Road and Fortess Road.

Finally the 'hotspots' of British non-responders living in group F (Figure 60, c) are found in the southern central part of the Borough. Hotspots are found along Maiden Road and the rail track in Gospel Oak. The predominant hotspot appears to be bounded by Pratt Street in the north, Euston Road in the south and Eversholt Street leading to the southern section of Camden High Street in the west and Pancras Road in the east.

The density maps have not been standardised for population and therefore each hotpot is likely to be the result of the population density of the neighbourhood; they are the result of the absolute numbers of people living in certain areas. This was done deliberately because it enables the identification of geographical neighbourhoods where the largest proportions of women who did not respond to screening live. For social marketing campaigns this means targeting can be centred on the red hotspots in Figure 60. This should maximise the efficiency of reaching the greatest number of people in the relevant cohort (segment), consequently this should improve the effectiveness of any campaigns for lowering the requirement resources and effort.

Whilst traditionally age and ethnicity are the traditional discriminators in this type of analysis, using a geodemographic classification crossed with a cultural, ethnic and language indicator the results provide a richer and insightful picture of non-responders. This combined with a cartographic representation of the geographic distribution of these different population sub-groups of women provides a useful way of gathering evidence to support specific targeted campaigns in different spatial locations.

7.4.3.4 Practice profiles

Using the technique developed first in the diabetes case study and then evaluated in the practice profiling case study, it was possible to identify general practices that have greater than expected likelihood of non-response to screening invitations. This technique will enable the spatial delineation of expected uptake after taking into account the variability of screening uptake across different population sub-groups.

In this step, expected uptake rates are calculated using the same sample of non-respondents. To calculate actual observed patterns of non-response to breast screening by cross referencing the neighbourhood Group for the 46 practices in Camden would yield very small numbers and render the results invalid. To remove the unreliability introduced by the small numbers that arise from dissecting the dataset to form a number of cross-tabulated variables, in this case neighbourhood Group and general practice, the expected uptake for general practices is calculated.

For each practice, the percentage of registered patients per Mosaic Group was calculated. These percentages were then multiplied by the previously calculated index value for each neighbourhood Group. The final step involved summing the results for each practice to yield an overall index value, highlighting the likelihood of women in a given practice attending breast screening.

Figure 61 shows the differences in predicted response rates for general practices across Camden. The map was classified using quintiles. Practices coloured in dark red/brown with the largest circles are representative of practices where the expected rate of non-response was greater than the Camden average. Therefore given the demography of the patient mix these practices are more likely in future years to have below average screening response rates. Those general practices in the south of the Borough tend toward having above average response rates. This means that by accounting for the differences in neighbourhood characteristics of the patients registered with general practices, it is possible to predict the likely response to breast screening calls for the registered eligible patients.

Six of the eight practices in the southern wards are expected to have non-response rates marginally higher than the other practices in Camden, but two practices stand out because they have unusually high expected scores of non-response compared to the other GP practices in Camden. These two practices are labelled in the diagram below. Practice F83672 with an index score of 125 is expected to have over 25% more patients not responding to calls for breast screening, considerably above the average for Camden which was set at 100. Also practice F83043 is predicted to have 5% more patients than the Camden average who do not attend screening.

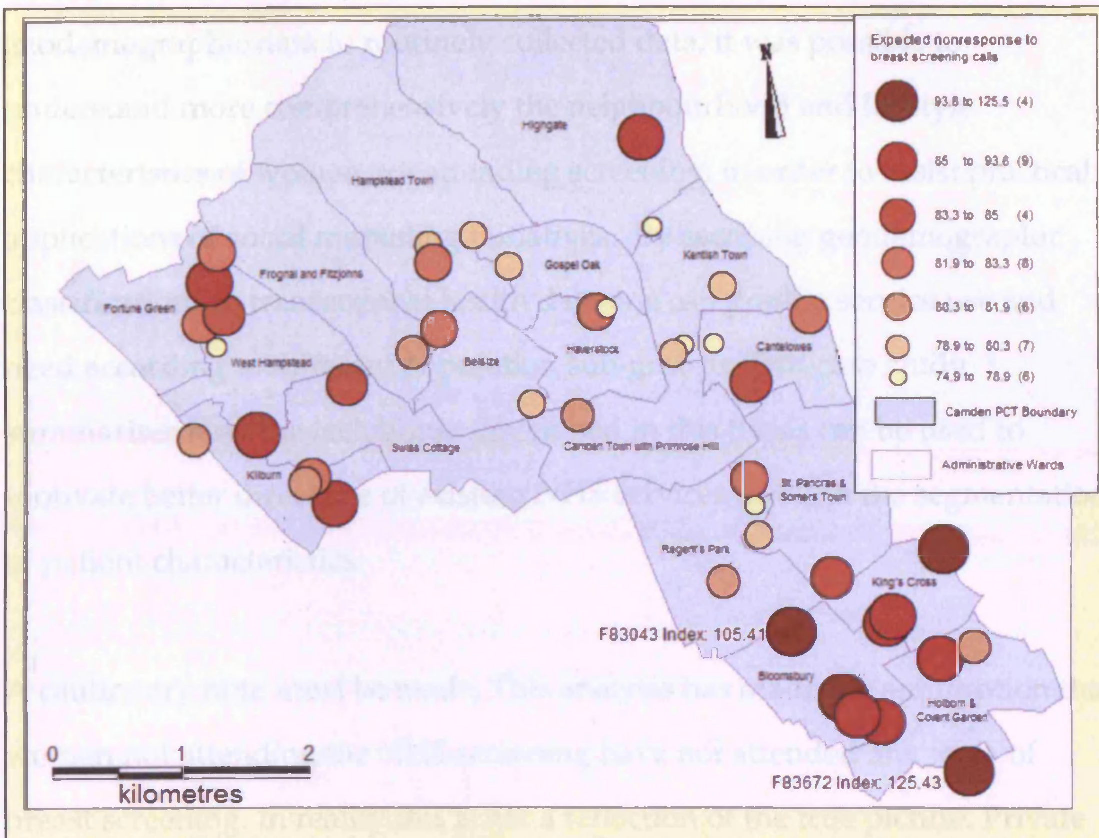


Figure 61 : Predicted rates of non-response to calls for breast screening by general practices in Camden

In the context of social marketing, this map has the potential to be a useful tool. Accounting for the differences in population characteristics of GP patient registers, in the south of the Borough all but one of the practices has higher expected scores of non-response in Camden. The exception is the practice linked to the university (UCL) where the patient cohort is younger than that of the screening cohort and has a very low risk of non-response. This cluster of practices, with higher than expected risk, might be the focus of group centred social marketing initiatives.

7.4.4 Discussion

This case study has set out to use neighbourhood (geodemographic) techniques to provide an enriched picture about the women who do not respond to breast screening invites in Camden. Through the application of geodemographic data to routinely collected data, it was possible to understand more comprehensively the neighbourhood and lifestyle characteristics of women not attending screening, in order to assist practical applications of social marketing initiatives. By ascribing geodemographic classifications to transactional health data one can predict service use and need according to different population sub-groups. This case study summarises how the techniques developed in this thesis can be used to motivate better direct use of existing NHS services through the segmentation of patient characteristics.

A cautionary note must be made. This analysis has made the assumption that women not attending the NHS screening have not attended any form of breast screening. In reality this is not a reflection of the true picture. Private clinics offer screening services to women who can afford to pay or who have private insurance as part of their benefits package at their place of work, but these data are not available to NHS practitioners. If this is the case, then in Camden it would be a fair assumption to say that women living in the neighbourhood Group A are more likely to attend private screening. Consequently the index values for Group A are likely to be considerably lower.

Recent studies have shown that affluent women are more likely to develop breast cancer, because they are more likely to exhibit the risk factors that are associated with breast screening (Banks et al., 2002). But women from lower socio-economic population sub-groups tend to have higher mortality and morbidity rates linked to the disease because they are more likely to present to the doctor with more developed stages of breast cancer (because they are not attending routine screening). In order to improve the equity of the service, women from neighbourhood Groups F should be encouraged to present for screening.

The creation of an aggregated index value for Camden general practices has two uses. Firstly, it enables general practices to be compared, whist having accounted for the socio-economic differences of each registered practice population. Secondly, it enables the prioritisation of campaigns and health promotions to practices with the highest index values. In this case the practices in the south of the Borough are likely to have high non-response rates for women not attending screening. It would be worth considering closer working with these practices to encourage the attendance of women at screening (for example: encouraging opportunistic discussion with female patients when they present at the surgery).

If we return to the first map presented in the breast screening case study (Figure 55) which was a simple density map of all non-responders for the age-cohort, it provided very little useful information from which social marketing campaigns could be derived. Combing the dataset with the geodemographic classification and combined with a derived classification of ethnic origin (section 7.4.3), did provide a more useful cartographic representation of the population distributions with the greatest need to practically inform health professionals about segmenting their cohorts. Furthermore, despite the little differentiation that was observed in risk index across neighbourhood Groups (section 7.4.3.2), the geodemographic analysis enabled the identification of residential neighbourhoods where women are less likely to attend screening and who are more likely to present with more advanced stages of cancer and therefore suffer greater morbidity. This analysis provides a useful benchmark for enabling health professionals to reduce morbidity by specifically targeting these neighbourhoods. Using simply age-specific cohort data would not reveal this. In the instance where geodemographics, may not be the most effective discriminator (as the analysis for neighbourhood Groups highlighted), they still have a place as a tool for equalising response rates across the risk groups.

The overall objective of this chapter was to substantiate the use of a geodemographic framework and its application to operational health data. The case study examples have gone some way to evaluate the ability of geodemographic classifications in exploring where geodemographics can contribute to targeting populations in social marketing and exploring response rates to health services and by corollary contribute to strategies aimed at equalising these response rates across different types of communities.

Chapter 8

Summary and discussion of analysis

8 Summary and discussion of analysis chapters

In the previous chapters, the geodemographic framework for measuring health outcomes was first developed in Chapter 6 using application areas linked to health harming behavioural choices: smoking, drinking and obesity. This chapter illustrated how useful geodemographics can be for exploring local differences in lifestyle behaviours, by enhancing their usefulness through the addition of national health surveys. It highlighted that smoking behaviour was highly differentiated across different social Types, but that obesity is less differentiated and appears to have a more even distribution across society. The analysis demonstrated that obesity was not endemic in only one particular social group and traversed the social spectrum. This framework was then substantiated in Chapter 7 using case studies related to diabetes and breast screening whereby operational health data were used to add-value and local level detail. Within the diabetes case study efforts were made to assess the scalability and validity of the measures within the context of two types of scales: formal (which corresponds to the organisational units of the NHS) and functional (that correspond to the local neighbourhoods). The results implied measures could be successfully scaled to facilitate prediction of diabetes risk across population sub-groups. These sizable chapters ended with a final case study that drew together all of the conceptualisations and applications with a view to exploring the extension for understanding real world issues linked to screening uptake. In this section response to breast screening invitations were assessed to inform public health practioners of possible population sub-groups that are not attending.

During the various inductive analysis, a number of recurring themes and points of discussion were identified and reiterated. It is these themes that are considered in this chapter, drawing together the analysis into a detailed discussion. Firstly, it was demonstrated in Chapters 6 and 7 that a geodemographics classification enhanced with health data was able to discriminate various patterns of health-harming lifestyle behaviours and diseases of comfort. Geodemographic Types for local neighbourhoods successfully differentiated patterns of smoking, obesity, diabetes and teenage pregnancies to name just a few. This raises the question of the validity of geodemographics as generalised representations of social reality. Do geodemographic classifications provide adequate representations of real life neighbourhoods? Secondly, standard deviations were used throughout Chapters 6 and 7 to explore the distributions of the indicators across the neighbourhood Types, whereby large standard deviations indicated a wider distribution and a greater potential for the variables and their associated measures to be useful in a social marketing campaign targeted at the neighbourhood. This is because a large distribution implies that there is greater differentiation across neighbourhood, and thus geodemographics are an effective discriminator. This led to the speculation about the ability of the classifications to be effective discriminators. For different neighbourhoods, can geodemographics effectively discriminate differential patterns of disease or lifestyle risk? The third recurrent theme that was evident in the analysis was the notion that geodemographic classifications are a practical and useful alternative to standard deprivation measures. Geodemographics have the potential to provide a more detailed description of the homogeneity and heterogeneity of local areas, so are geodemographics a viable alternative to deprivation measures? The fourth and final theme continuing across all of the case studies is the idea that the classifications could potentially provide a practical mechanism for exploring social space and health status based upon the social similarity and health outcome similarity of neighbourhoods. This theme arose from the exploration of composite surfaces for measuring health behaviours in Chapter 6, whereby it was evident that variables could be combined to consider health behaviours as a whole. Geodemographic classifications that are ascribed with health data provide an enriched picture of the local neighbourhood and by corollary the social space of the residents. Is it possible to take some of the notions of social capital that combine various capital elements of economic, knowledge and social (Bourdieu, 1997) to create social facts relating to neighbourhood spaces (derived from geodemographic classifications) together with a combined indicator of health status to explore these in a multi-dimensional space, comprised of social, geographic and health?

In the first section of this chapter, the first question to be explored is "Do geodemographic classifications provide adequate representations of real life neighbourhoods?" This is discussed with respect to the distribution of social class and neighbourhood Types as classified by respondents of the Health Survey for England (HSE). A standard dissimilarity index is used to consider the homogeneity within each individual neighbourhood Type to determine if they are useful and provide an adequate descriptor of the population representation of neighbourhood scale.

The second section considers the question "can geodemographics effectively discriminate differential patterns of disease or lifestyle risk?" It takes an idea presented by Webber (2004) to create total weighted deviation (TWD) scores of the health measures created throughout Chapters 6 and 7. The TWD scores are used to identify which health outcomes are most appropriately discriminated by geodemographic Types, and as a consequence are most likely to be receptive to social marketing campaigns.

The penultimate section of the chapter investigates the question "Do geodemographics present a viable alternative to deprivation measures because of their improved ability to provide effectively discriminators across a relevant spectrum of the population?" In this discussion the health deprivation domain of the Index of Multiple Deprivation (2004) is applied to a geodemographic classification in order to explore how useful they are at distinguishing health deprivation (a commonly used population discriminator), to determine its suitability for informing social marketing. The fourth and final section of the chapter investigates whether "Geodemographics can aid the visualisation of multi-dimensional space?" The section reviews the debate surrounding social and geographical scale with respect to organisation hierarchies that relate to formal aspatial administrative units (organisational) scale and functional (local spatial neighbourhood) scales approximating the sociological construct of the neighbourhood. The themes are synthesised together in this final section to create a visualisation of social and geographical space of Camden neighbourhoods.

8.1 Geodemographics: representations of reality?

Returning to the models of social determinants of health documented in section 2.3.2, it is possible to consider each one of us as a unique combination of hereditary characteristics and susceptibilities together with influences from our surrounding living environments (past and present), cultural perspectives and practices, and societal pressures and normalities. Despite these individual differences, these factors interweave together in various ways to contribute to both our health as individuals and the overall population health outcomes (Smith 2006 and Graham, 2001). Invariably, as suggested by the results of the case studies in Chapters 6 and 7 some characteristics and health outcomes are shared across population sub-groups.

Thus whilst we experience ourselves as individuals, it is important to understand that people are often grouped or structured into social categories (Smith 2006). This implies social similarity can be determined as a result of the processes of social clustering. In the past, sociological thinking grouped sub-groups of the population together according to economics and culture by exploring the differentials in income and status either following the thinking of Marx or Weber. The renaissance of geodemographic classifications in the last decade as acknowledged by Longley (2005) represented a shift in sociological thinking moving away from these fundamental notions of classification (Burrows and Gane, 2006). Previously, people were grouped together by income and status to create a measure of social space; comprised of social, economic and cultural capital. It could be considered that geodemographics sort people and places by joining social spaces with geographical spaces in order to develop predictors of consumption, values and tastes. In a return to a variation of social area analysis, geodemographics provides this view of socially similar neighbourhoods (Harris et al., 2005).

Geodemographics present a straightforward method of conceptualising the interactions of socially similar people by arranging social divisions of populations in accordance with their spatial arrangements. For example, elderly people who have difficulty walking may try to relocate and retire to bungalows in seaside locations with good amenities and access to medical services. Thus, people in the same predicament move to the area and socially cluster in neighbourhoods with comparable characteristics. It can be argued that lifestyle together with uniqueness and type of place provides an important determinant in choice of residence.

These representations of social similarity consider space as the medium in which fixed social processes occur (Smith, 2006 on Simmel, 1855) and are reflected in the neighbourhood Types that comprise the classification. The geodemographic classification used in this thesis identifies 61 different social divisions, reflecting the 'average characteristics' of the population associated with their spatial arrangement. One underlying assumption of these methods of social stratification which is critical for their application in the public sector is whether or not they can be demonstrated to provide acceptable representations of reality.

To test their validity as representations of reality it is necessary to return to the Health Survey for England (HSE), which was used in Chapter 6 to develop a range of indicators related to lifestyle behaviours. The Health Survey for England not only surveys the population of England about their health outcomes but asks questions related to conventional methods of social stratifications based on social class and income. Each respondent is asked to give details about their occupation and these occupations are then grouped together accordingly to return a classification of socio-economic occupation. One would expect that if the geodemographic neighbourhood Types are indeed reflective of the average representations of the social similarity within survey respondents then the occupational classification for each neighbourhood type should be consistent, homogenous and certainly should not contain a wide range of diverse occupations.

Figure 62 outlines the distribution of HSE responders (as used in Chapter 6) and their corresponding socio-economic classification cross-tabulated by neighbourhood Type. Counts for respondents for six different occupational classes are summarised for each neighbourhood: unskilled manual, semi-skilled manual, skilled manual, skilled non-manual, managerial technical and professional occupations each classified with a different colour. The neighbourhoods intuitively linked to those associated with traditional notions of deprived, low income, public housing, multi-cultural neighbourhoods have greater numbers of people working in unskilled and semi-skilled manual occupations (labelled A). Socially categorised neighbourhoods closely linked to affluent high achievers also appear to have a predominance of professional and managerial technical occupations.

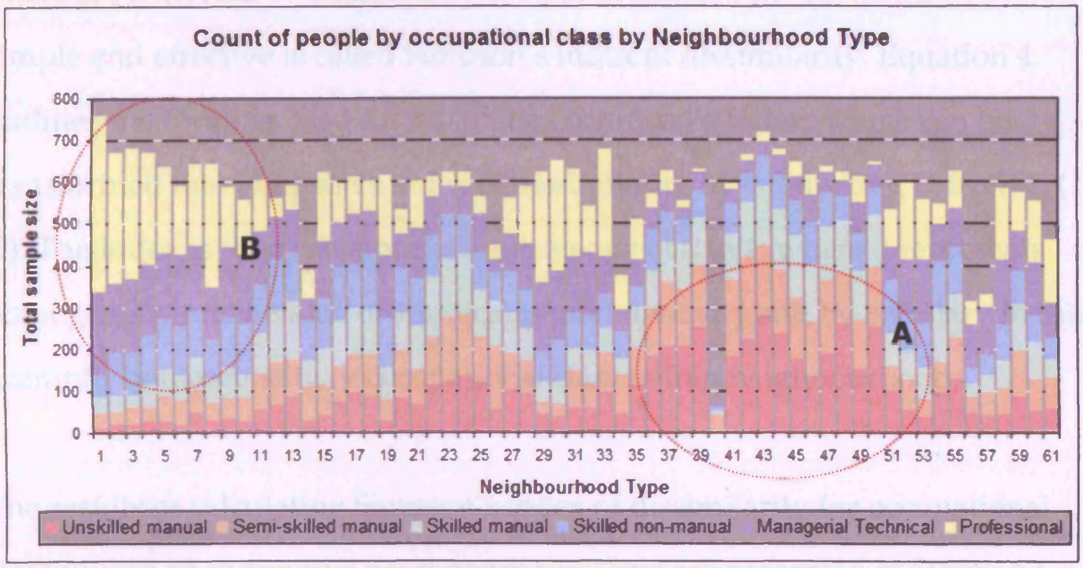


Figure 62: Counts of health survey for England responders and their corresponding socio-economic classification cross tabulated by neighbourhood Type

It is possible to test the measure of heterogeneity and/or homogeneity of occupational classes within each neighbourhood Type, using standard dissimilarity indices. Dissimilarity indices are utilised in ecological studies to explore species diversity, whereby diversity indices are calculated to compare ecological communities. They have also been used by demographers to explore ethnic diversity and segregation (Reardon, 2006). To assess the diversity of occupations within neighbourhood Types, two elements must be considered; richness (the number of responders in a sample) and evenness (the number of responders for each occupation). If the distribution of different occupations within a neighbourhood Type is diverse then there will be an equal number of people within each occupational class for each of the

61 neighbourhood Types. The priori expectation is that the neighbourhoods in each Type will be socially similar and have low values of dissimilarity.

$$D = \frac{\sum n_t(n_t - 1)}{N(N-1)}$$

D = Simpson's Index
N= total number of all survey responders
n_t= total number of all survey responders within an occupational class

Equation 4

There are a myriad of dissimilarity indices available but one that is both simple and effective is called Simpson's index of dissimilarity. Equation 4, outlines the formula used for calculating Simpson's Index, which can be transformed into Simpson's index of dissimilarity by subtracting D from 1, (1-D). The index is then transformed onto a scale of 0 to 1, where 0 represents homogeneity and a value of 1 indicates total diversity and by corollary in this example, heterogeneity of occupational class within neighbourhoods.

The results of calculating Simpson's index of dissimilarity for occupational class within an individual neighbourhood Type are presented in Figure 63. The only neighbourhood Type which can be considered homogenous from the perspective of occupational class is Type 14; a neighbourhood representing military quarters. This Type has an index value of 0.08 for dissimilarity. Indeed, 32 of the neighbourhood Types have scores greater than 0.8 and all but one Type has a dissimilarity score of greater than 0.6

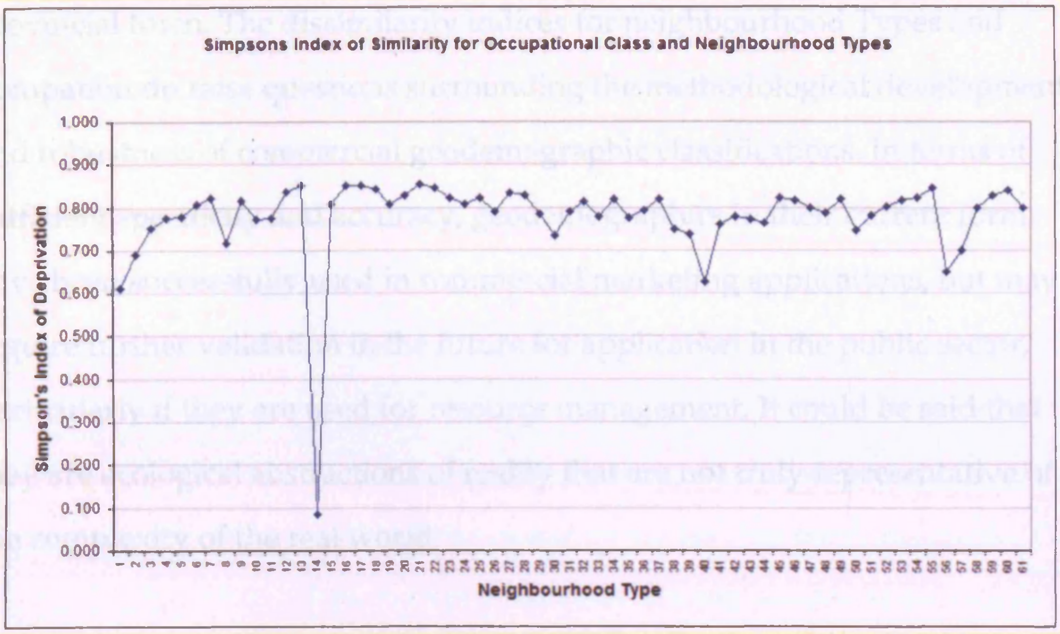


Figure 63: Simpson's index of dissimilarity for occupational class and neighbourhood Type

At first glance the results of Simpson's index of dissimilarity appear to contradict the notion that individual geodemographic Types are relatively homogenous and are representative of socially similar population sub-groups. The only Type that is actually approaching homogeneity is Type 14, which corresponds to people living in military quarters. It seems logical that this type is homogenous as only military employees and their families can live in such places and cannot be expected to be too heterogeneous. Therefore, it is relevant at this point to reiterate the concept that such classifications are merely indicators of the average characteristics of neighbourhoods. The notion that social conditions can be summarised in a single univariate measure is difficult, since it is difficult for a single measure to encompass family structure, employment, status, demography, ethnicity and income etc. The socio-economic classification of occupation is not without its limitations and indeed the results could simply be an artefact of the method of data collection and classification of occupation itself. There are different connotations of occupation that will vary spatially according to place. A solicitor in the city of London is very different to a solicitor is a provincial town. The dissimilarity indices for neighbourhood Types and occupation do raise questions surrounding the methodological developments and robustness of commercial geodemographic classifications. In terms of sufficient specificity and accuracy, geodemographics in their current form, have been successfully used in commercial marketing applications, but may require further validation in the future for application in the public sector, particularly if they are used for resource management. It could be said that they are ecological abstractions of reality that are not truly representative of the complexity of the real world.

8.2 Geodemographics: effective discriminators?

Whilst the uncertainties of methodological robustness and accuracy of the classifications as reflections of realities are difficult to determine, we have seen during the development of health indicators, that despite these difficulties, geodemographics application have a place for facilitating small area urban analysis for social marketing initiatives. Geodemographics systematically and logically sort people and places (Burrows and Gane, 2006). The classifications enable social spaces to be connected to geographical spaces in order to develop predictors of consumption, values and tastes, and, as was seen, to assist in describing socio-spatial differences in health outcomes. In many respects, this form of analysis and categorisation marks a return to the Shevky and Bell approach to social area analysis of the 1950s, but with notable differences. Geodemographics, in their nature are avowedly inductive, whereas Shevky and Bell postulated the existence of three or four latent dimensions underpinning the data which were included in classification. In practice, geodemographics have the potential to be deductive if the analyst were able to change and configure the weighting scheme to match the specific purposes of each application.

The residential segregations evident from the spatially distributed geodemographic Types in Camden reflect the social differentiation of society in a geographical space. For an application to social marketing, the extent of population heterogeneity in Camden should be appraised. This will determine whether or not there is sufficient diversity to warrant contrasting health marketing campaigns. Once again Simpson's index and Simpson's index of dissimilarity (see previous section) were calculated for the population distribution of neighbourhood Types for Camden's residents. The resultant index of dissimilarity for Camden was 0.803. As this number is close to 1, the results indicate that neighbourhood diversity in Camden is high and is comprised of very varied communities. The results for Simpson's index of dissimilarity of neighbourhood Types in Camden reflects the multi-dimensional nature of its residential communities, and indicates that specific social marketing campaigns would be appropriate for different sub-sections of the population. The next question is to determine what types of diseases/conditions and lifestyles from the case study results would be most suitable for targeting with campaigns.

In the analysis conducted in Chapters 6 and 7, an array of expected health outcome variables was predicted for a range of illnesses and conditions. They highlighted, to varying degrees, the differentiation of outcomes across the spectrum of neighbourhoods. Many of the exploratory analysis conducted were concerned with identifying the existence of patterns across these neighbourhoods. One of the four key principles of social marketing is place, the others being product, price and promotion. A thorough understanding of the place aspect of social marketing enables development of the remaining principles within a campaign of public health intervention. But whilst the developed indicators provide an enriched picture about both the place and the product (the type of illness, disease or condition) the vast array of variables present health professionals and social marketers with a dilemma as to how priorities might be identified and acted upon?

In marketing, Webber (2004) suggests this can be appraised by using an index based upon a total weighted deviation (TWD) measure. This calculation is used to examine the degree to which an indicator variable is an efficient discriminator. In this example TWD was used to determine which of the indicators developed in Chapter 7 perform better across the neighbourhood classification, identifying the conditions/illnesses which would benefit the most from social marketing campaigns in Camden targeted at neighbourhoods.

Total weighted deviations for different conditions and illnesses were calculated for the index values developed in Chapter 7 (bearing in mind the limitation that the values have a lower limit close to 0, but no upper limit). For each neighbourhood Type the index scores (Ds) for each condition were subtracted from the average score (100). This revealed the extent to which the particular neighbourhood deviates from the average neighbourhood. The values were squared then square rooted to remove negative values, returning the absolute deviation of the neighbourhood Type from the average neighbourhood. These results were then multiplied by the proportion of the population (p) living in each neighbourhood Type (n) in Camden. The resultant values for each neighbourhood Type are then summed to produce the total weighted deviation. Using diabetes type II as an example, for neighbourhood Type 01, which comprises 22% of Camden's total resident population, the total deviation for neighbourhood Type 01 is calculated using the following method: [(sqrt((100-18)^2))* 0.22], equation 5.

$$TWD = \sum^{n=21}\left(\sqrt{(100-D_s)^2}\right)pn \qquad \text{Equation 5}$$

D_s= Disease/condition index score

p = percentage of population

n = neighbourhood Type

The total weighted deviation was calculated for ten of the different diseases/conditions used in Chapter 7. The results are outlined in the graph in Figure 64. The graph is organised in descending order, with the diseases/conditions with the highest TWD found on the left hand side of the graph. Given the population distribution of Camden's residents according to their neighbourhood Type, social marketing campaigns could be a suitably prioritised for interventions according to their TWD score. The diseases/conditions ranked with the 5 largest TWD scores in Camden are: alcohol and drug abuse, mental health issues, teenage pregnancies, diabetes type II and asthma.

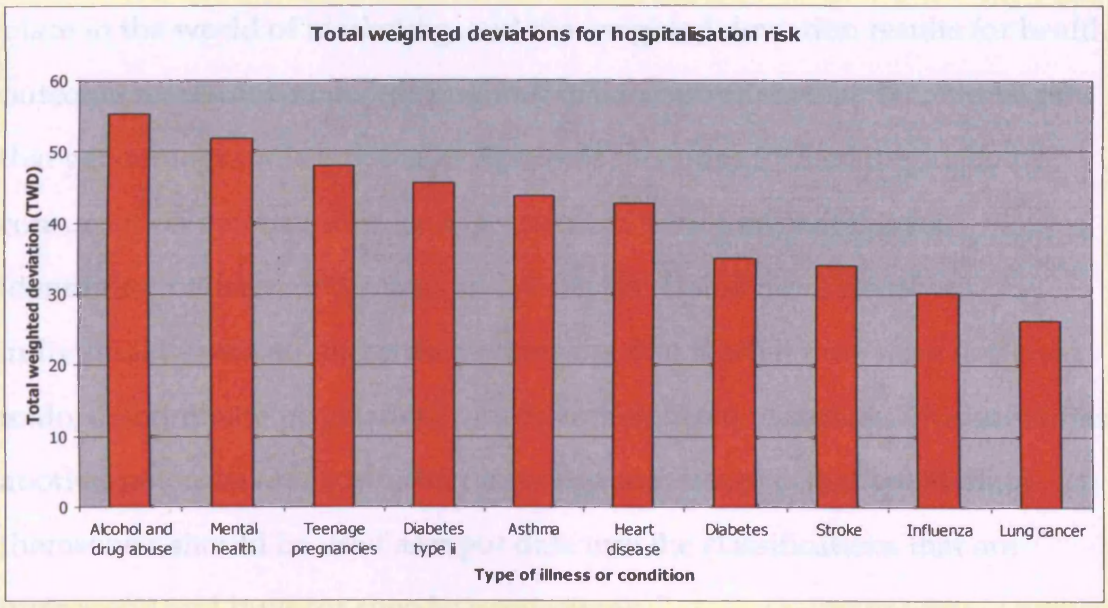


Figure 64: Total weighted deviation scores for hospital risk of 10 conditions/diseases for Camden's population

The results of these calculations indicate the diseases and/or conditions that social marketers should consider for campaigns in Camden. The TWD calculations indicate that for diseases of comfort the index scores vary considerably across the neighbourhoods and marketers should achieve high efficiency if targeting populations for the disease listed. As to be expected, diseases and conditions associated closely with lifestyle choices and behaviours are better discriminated by the Mosaic geodemographic classification. This reaffirms the appropriateness of using geodemographic classifications because of their discriminatory ability to characterise socially dissimilar neighbourhoods and their corresponding health requirements. Neighbourhood classifications are less appropriate for describing and selecting intervention campaigns for disease where lifestyle choices are less likely to be the predominant cause. This identifies a limitation of these types of classifications. Contemporary geodemographic classifications as we know them today have their origins in the measurement and identification of areas of deprivation and date back to the Liverpool social malaise study (Liverpool City Council, 1969). Despite this, their formative years and development took place in the world of marketing, and the weighted deviation results for health outcome scores developed throughout this thesis reflect that. It could be said that geodemographic typologies represent identities formed originally by consumption patterns. It is for this reason that they are suitable for identifying patterns of diseases and lifestyles related more closely to individual behaviour and attitudes because that is what they were designed to do, discriminate populations based on their lifestyle choices. This identifies another potential area for further investigation. It seems that health data themselves should be used as input data into the classifications that are purposeful and built for specific applications.

8.3 Geodemographics: an alternative to deprivation measures?

A result seen many times in this thesis was the apparent advantages gained by using geodemographic classifications for analysis as compared to more traditional approaches presented by deprivation measures. The previous section identified the usefulness of geodemographic classifications to discriminate neighbourhoods based on lifestyle diseases and conditions. In the various case studies presented in Chapters 6 and 7, geodemographic (neighbourhood) classifications enabled the development of health indicators to be applied to small areas representing, on average, 16 households. The creation of indicators at this spatial scale ensured that different analysis distinguished relevant and specific areas with the greatest observable health needs and requirements.

The importance of the indicators is reflected in their relevance to social marketing. Market research techniques are regularly used by companies in the private sector to target their products and services at particular population sub-groups of consumers. However, since the mid 1990s there has been a paradigm shift in public policy, in part because of the pressure to demonstrate value for money. Public sector organisations are now slowly adopting these traditionally commercial techniques in order to gain similar insight into the differential health needs of resident populations.

In 2007 the total budget for the NHS was set at £90bn and is predicted to rise to £110bn in 2010 (HM Treasury, 2007). Recent policy handed down power and responsibility to local PCTs who control their own portion of this budget. Their budget is to be used to develop and provide services that will meet the health needs of the local population as dictated by government policy, for example the public health white paper (Department of Health, 2004). There is growing evidence for the contribution that patient-focused social marketing can make to improve the impact and effectiveness of behavioural interventions, be it in policy formulation, strategy development or implementation and delivery (National Social Marketing Centre, 2007). But to ensure social marketing is successful appropriate tools must be in place for identifying health need.

In both models of social determinants of health presented in section 2.3.2, it was evident that many compositional and contextual factors interplay, giving rise to poor health outcomes and differential health needs, one of which is income and its association with deprivation. But whilst deprivation is a commonly used discriminator of health need (section 3.1), we must remember that, as a term, deprivation itself is a controversial and elusive concept (Boyle et al., 2002), describing the notion of accessibility to material necessities such as adequate food & heating as well as social facilities affecting individuals, households, families, groups and institutions (McCulloch, 2001; Hirschfield et al., 1995).

In Chapter 3, section 3.1, a review of the most commonly used deprivation measures in the UK was undertaken. The most recent and frequently used deprivation measure in England was/is the Index of Multiple Deprivation from 2004/2007 (IMD 2004; 2007). It is a composite weighted deprivation index combining 37 measures across 7 domains: income, employment, health and disability, education, skills and training, barriers to housing and services, living environment and crime (ODPM, 2004). It has become commonplace for health and other public sector professionals to use deprivation measures as proxy indicators for locating geographical areas with predicted poor health outcomes amongst the residential population. But as seen throughout Chapters 6 and 7, it is not always a suitable health discriminator because of the variables used to develop the indicators, the areal unit at which it is measured and the timeliness of the input data.

The IMD 2004 contains a health domain which utilises 4 distinct datasets to create a deprivation index score specifically for health: mortality rates – years of potential life loss (YPLL), comparative illness and disability ratio, emergency admissions to hospital (derived from Hospital Episode Statistics) and adults under 60 suffering from mood or anxiety disorders (based on prescriptions, hospital episode statistics, suicides and health benefits data).

Figure 65 below highlights the distribution of the super output areas (SOA), (defined in section 5.3.1) for Camden according to their scoring in the IMD 2004 health deprivation domain. The map has been classified using quartile ranges to show the 25% most deprived SOAs in Camden, coloured in dark brown on the image. With the 25% least deprived, in terms of health, highlighted in light cream.

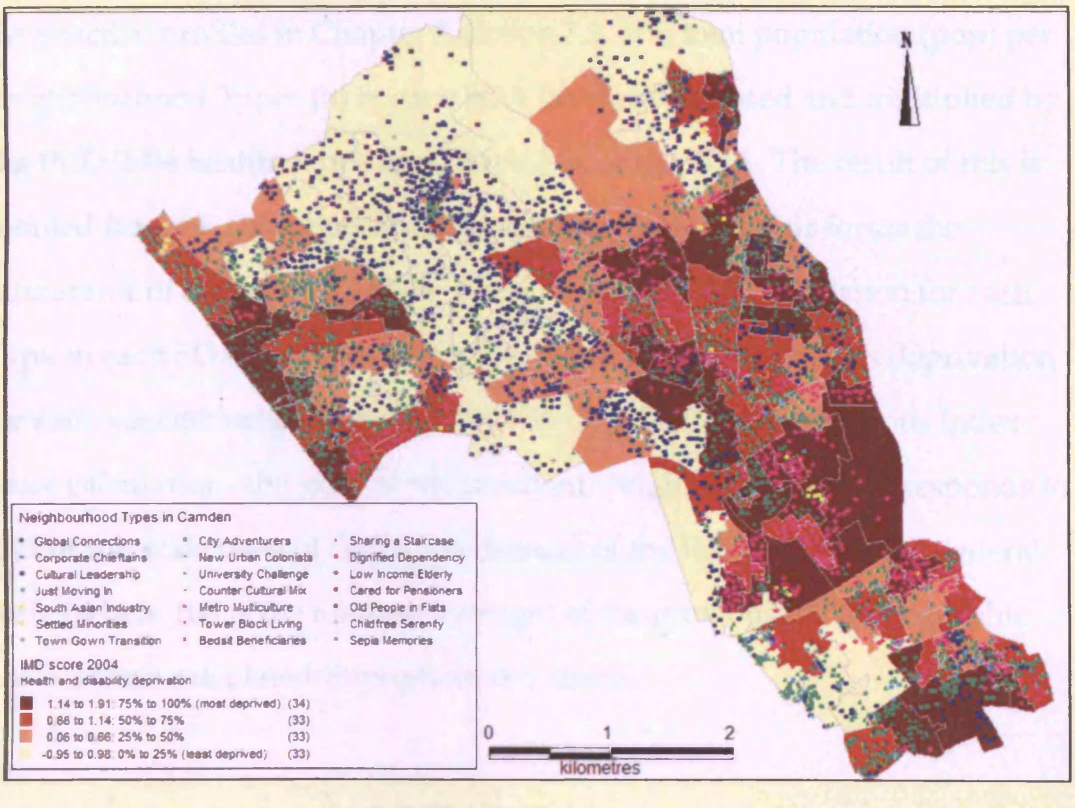


Figure 65: Map of health and disability deprivation for Camden super output areas with neighbourhood Types

The individual points on the map represent the centroids of the unit postcodes and their corresponding neighbourhood Types for Camden, with each Type classified by a different shape and colour. At the scale for which the map has been produced it is difficult to identify the distinctive shapes (of the points) but the colours represent the hierarchical groupings of the neighbourhoods, and so should be used for interpretation in this instance, to highlight postcodes belonging to the same neighbourhood Groups. The map visually illustrates the heterogeneity of neighbourhood Types within the individual SOAs, and introduces the important notion of scale (discussed in detail in following section). The 25% most health deprived areas coloured in dark brown do not adequately capture the true extent of health inequalities and variations in health outcomes as observed in Chapters 6 and 7.

The health deprivation scores from the IMD 2004 were linked to the geodemographic classification, adopting a similar technique as used to create the practice profiles in Chapter 7 section 7.3. The total population (pop) per neighbourhood Types (n) in each SOA (a) was calculated and multiplied by the IMD 2004 health deprivation score (d) for the SOA. The result of this is totalled for each neighbourhood (n) type per SOA (a). This forms the numerator of equation 6. This is then divided by total population for each Type in each SOA to return a weighted average index of health deprivation for each specific neighbourhood Type in Camden. Unlike previous index score calculations the scale of the resultant weighted average corresponds to that of the scale used of the health domain of the IMD, and not the general scale (where 100 is the national average) of the previous geodemographic index scores calculated throughout this thesis.

$$Index_n = \left(\frac{\sum d_a \, pop_{an}}{\sum_n pop_{an}} \right)$$

d = health deprivation score
pop = population
a = Super output area
n = neighbourhood Type

Equation 6

Figure 66 highlights the results from the calculation of equation 6. The weighted average health deprivation index of for each specific neighbourhood Type shows which Types are more likely to be subjected to health deprivation as defined by the IMD 2004 health and disability domain. According to the results in Figure 66 the higher the score the more 'health deprived' the neighbourhood, indicating that a neighbourhood is much more likely to be subjected to negative health outcomes. Figure 66, indicates that neighbourhood Types 26, 40 and 49 are the most health deprived in Camden.

Note, this graph is not on the same scale as the previous index graphs used in this thesis, and negative values result because the methodology used to create the IMD health domain scores contained negative values.

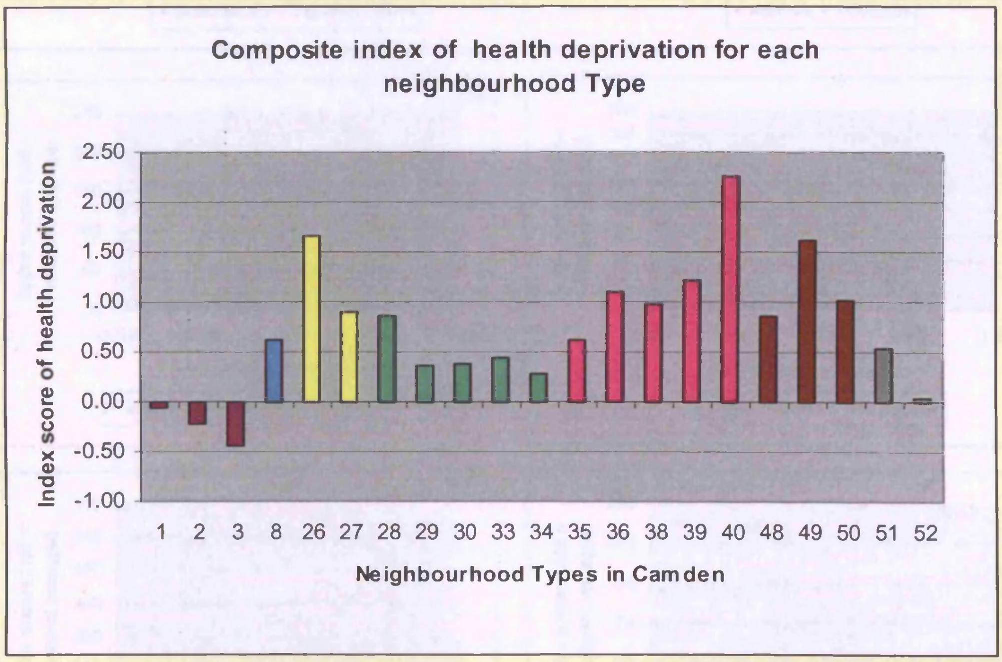


Figure 66: Health deprivation scores for Camden neighbourhoods

Whilst this is useful, it has little relevant application for social marketing. If neighbourhoods were targeted according to this deprivation score alone, it is likely the effectiveness of the campaigns would be greatly reduced because it does not adequately indicate health variation for different conditions. To prove this, if health deprivation alone were a suitable indicator for discriminating health need, one would expect a linear correlation between its measure and the various calculated health indicators. As a consequence the health deprivation scores have been plotted against a number of the indicators developed for different diseases of comfort and health conditions (from Chapters 6 and 7), and as shown in Figure 67.

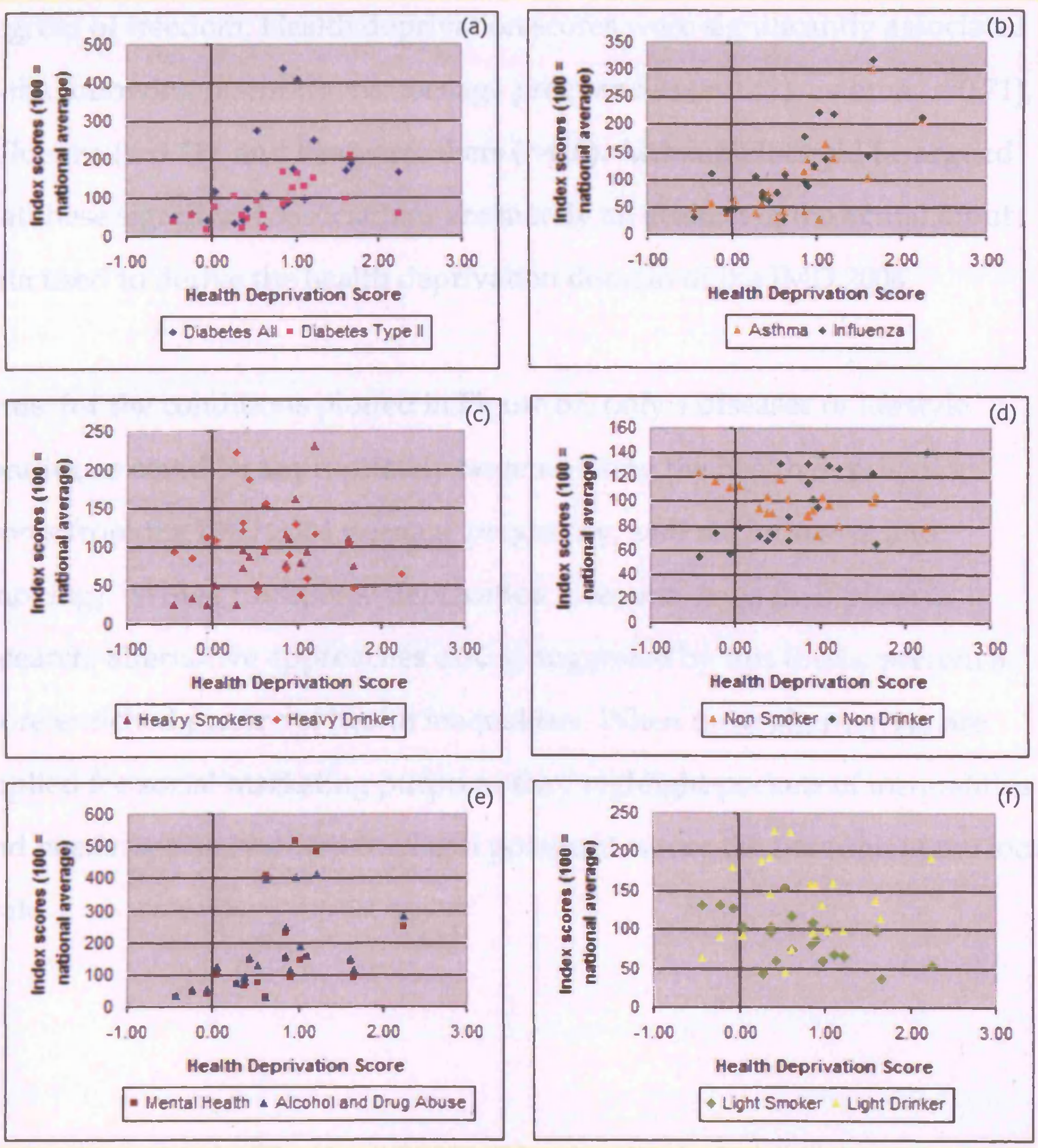


Figure 67: Graphs of health deprivation scores vs index scores for disease of comfort

The graphs in Figure 67 indicate that health deprivation scores for different neighbourhood Types rarely strongly correlate with neighbourhood distributions of diseases and conditions of comfort. The correlation coefficients returned only weak to moderate strengths of association with no coefficients being greater than 0.8. Statistical significance testing of the correlation coefficient for each of the diseases indicated that only a few conditions showed a significant likelihood, less than 1 in 100, of the patterns not occurring randomly; where the critical value was set 0.549 with 19 degrees of freedom. Health deprivation scores were significantly associated to the following distributions: teenage pregnancies (r=0.62), asthma (r=0.71), influenza (r=0.71), and heavy smokers (r=0.6). Although it could be argued that these significant associations are merely an artefact of the actual input data used to derive the health deprivation domain of the IMD 2004,

Thus, for the conditions plotted in Figure 67, only 4 diseases or lifestyle conditions could be appropriately targeted using the health deprivation scores from the IMD 2004 (teenage pregnancy, asthma, influenza and smoking). Whilst traditional deprivation measures have their place in research, alternative approaches can, as suggested by this thesis, present a more enriched picture of health inequalities. When these alternatives are applied for social marketing purposes they highlight pockets of inequalities and needs unevenly distributed and polarised across the borough at the local scale.

8.4 Geodemographics: identifying geographic and social space in the health domain

As we have seen in both of the previous chapters, the indicators developed in this thesis contribute to building a much more enriched picture of social groups in Camden for different neighbourhood Types. Health outcomes linked to lifestyle behaviour and choices appear to be successfully measured by combining national health surveys with local classifications of population sub-groups. This thesis suggests that the judicious selection of complementary data aids the development of indicators which contribute to the effective discrimination of the health outcomes most relevant to lifestyle and behaviour choices, which in turn are the most relevant for social targeting through social marketing campaigns.

This thesis has also highlighted how the more traditional quantitative approaches of social measurement, shied away from in the 1980s, when combined with new qualitative approaches can create tangible measures, for which sound targeted public decision making can be based. The literature review also noted that consideration of health inequalities must be made for both the geographical space and the social space which both influence the resultant indicators.

The connection between health outcomes and the social and geographical spaces in which they occur is undeniable. In both types of models of social determinants of health, the individually centric (Dahlgren and Whitehead, 1991) and the population centric health model (Friedman et al., 2002), location plays a crucial role in the identification of health inequality landscapes in the UK. Geodemographic measurement techniques assist in placing health outcome measurement within the realms of both of social and geographical space.

8.4.1 Geographical space and scale of lifestyles and health status

The essence of any geographical analysis is the central importance of location and place and scale. They both have considerable impact upon the study of health outcomes. The geographical aggregation of health data into administrative areas is now commonplace within the mandatory production of the annual public health report, produced by each PCT in England. Events, patterns and processes related to health must be explored according to their position in space and the context in which they are occurring. Indeed the role of geographical structures in space for health care is now commonplace in policy and decision making.

It is for these reasons that spatial analysis is concerned with problems that have an explicit spatial context, and that context on its own is not independent of surrounding situational data, i.e. the social space (de Smith et al., 2008). Many exploratory analysis are concerned with identifying patterns. Indeed, as has been previously mentioned, one of the four key principles of social marketing is place, the others: product, price and promotion. A thorough understanding of the place aspect of social marketing, whereby knowledge is built about the unique attributes and characteristics associated to a location, enables development of the remaining principles within a campaign or public health intervention. de Smith and colleagues note (2008, online), 'on its own spatial location is not interesting. The power of location comes not from location itself, but from the linkages or relationships that it establishes – from relative positions rather than absolute ones'. It is the attributes and features associated with a location that make that place interesting for research.

For health care policy and practice, knowledge of place and social space are key to effective targeting. The relative locations of neighbourhoods as places in health research provide insight and understanding of health related behaviours, diseases and other relevant phenomena. Indeed the attributes of these locations provide the key to unlocking information and knowledge about an area. The comparison of these attributes in close proximity to each other is another useful resource for exploring data patterns and relationships. To gain a true picture of health inequalities, the social and geographical space must be considered simultaneously. One can argue that geodemographics and GIS enable this to be realised.

A discussion of geographical space cannot be considered without the mention of geographic scale as it is a complex and important issue. Scale should be considered as a fundamental aspect of any location based spatial analysis, so prior to exploring social space, a review of geographical space and scale is relevant. As noted by Atkinson and Tate (2000), geographical scale provides the linkages between the distributions of properties and attributes in space and their representations. The limitations, complexities and fundamental problems created by scale in human geography have been extensively documented (for example: Goodchild and Proctor, 1997, Longley and Batty, 1996, Sheppard and McMaster, 2003 and Atkinson and Tate, 2000). The importance of scale was first presented in Chapter 3, and explored in the analysis Chapter 7, section 7.2 where the scalability of the indicators was appraised. The importance of scale for health data is irrefutable because scale influences many results, through the manifestation of issues related to both the ecological fallacy and the modifiable areal unit problem. It is essential to explore its effect on the formulation of a study, its role in information content, analysis methods, interpretations of findings and conclusion of patterns and processes.

Lam (2004, page 23) references four meanings of scale within geographic inquiry; spatial extent of study (geographical scale), data resolution (granularity of the data, also known as spatial resolution), extent of a spatial process and representation through a map (operational or analytical scale). It is not uncommon for the geographical scale of maps to be referred to as either large or small scale, but the actual meaning of these terms can be interpreted differently. In describing the same spatial data, a geographer may refer to it as small scale whereas an ecologist could refer to it as large scale. This is because, "the geographer is describing scale in reference to the ratio of map distance to earth distance whilst the ecologist is often describing scale in reference to the extent of the phenomenon studied or the spatial resolution that was incorporated in the study", (Jenerett and Wu, 2000 page 104). The terms large and small scale are sometimes counter intuitive and respectively relate to small and large entities (Montello, 2001). For example, human demographers and ecologists more often than not refer to small- scale issues corresponding to the local neighbourhood and are in fact referring to spatial extent. In this thesis and in health care policy the concept of local scale or small neighbourhoods refers back to the ecological interpretation of geographic scale.

The scale of analysis as defined by Montello is, "The size of the units in which phenomena are measured and the size of the units into which measurements are aggregated", (2000, page 13502). The ambiguity of analytical scale was noted by Tate and Atkinson (2000) who reiterated thoughts presented by Goodchild and Proctor (1997), that it can encompass both the amount of detail and the spatial extent of a geographic area. More commonly the scale of analysis is broken down into two components, summarised by Lam (2004) as the scale of measurement and scale of observation, and is often considered the scale of geographic understanding, ie the scale of knowledge acquirement (Montello, 2000). This conceptualisation of scale separates the process of acquiring knowledge into two halves: data collection and data analysis. It acknowledges that the scale for which data are observed and their ensuing data collection may and often will be different to the scale at which data are transformed into information by adopting a different measurement scale. Thus, recognising the two distinct requirement needs for data collection and data analysis, the division of scale into both measurement and observation was also adopted for the collection and production of the 2001 Census statistics (Rees et al, 2002). It is also evident in a lot of health datasets; often primary data are collected at the scale of the individual, but are then aggregated into different larger sized units in order to protect anonymity.

The units used in health data have two forms. They are either formal aspatial administrative units or functional areas approximating social space.
This means that the scale of geographical understanding adopted by health care organisations in the UK has two forms – the concepts of formal space and functional space as proposed by (Moon, 1990). If applied to Lam's notion of space (2004) both these forms of scale correspond to the measurement scale and the scale at which geographic inquiry occurs. The analysis throughout this thesis have considered and applied the measures to both of these forms of geographical spaces which are used commonly in health care.The choice of measurement scale and its corresponding form impact on the scale of geographical inquiry and the resulting knowledge acquired by the researcher.

8.4.1.1 Formal geographical space

Formal geographical space refers to the geographical partitioning of the United Kingdom into discrete adjacent spatial polygons. The structure of these polygons correlates to the organisational structure of the NHS. The spatial polygons demarcate the extent of responsibility and administration for different units within the organisation. In this sense they can be defined as formal geographical spaces because they identify a spatial boundary for which the organisation has a legal responsibility to treat and care for patients. In the case of the NHS, a number of formal geographies exist which nest inside each other, creating a hierarchy of organisational and formal geographical scales.

The structuring of both organisations and people into nested hierarchies emulates the hierarchical theory of nature, where nature is divided into a series of levels and abstractions which nest inside each other, characterising a hierarchical structure and interactions horizontally across levels (McMaster and Sheppard, 2004). The original manifestation of hierarchical theory defined the horizontal structure as a series of "holons". A holon is defined as, "a system that is whole in itself as well as within a larger system" (OED, 2002). The term is a derivative of two Greek words; "holos" (meaning whole) and "on" (meaning part). It is a form of organisation resembling pyramid, with each level subordinate to the one above it. The formal geography of the NHS is representative of such a structure, which is illustrated in Figure 68. Consequentially, it is this method of geographical organisation that determines the scaling of data and geography used in public health analysis and decision making and determines the constraints of geographical inquiry.

At the top level of the hierarchy, level 1, sits the Secretary of State for Health, who has ultimate responsibility for the NHS in England and is responsible to Parliament. The subordinate level is level 2, the Department of Health (DoH) and the NHS executive who are responsible for the defining strategic health care planning.

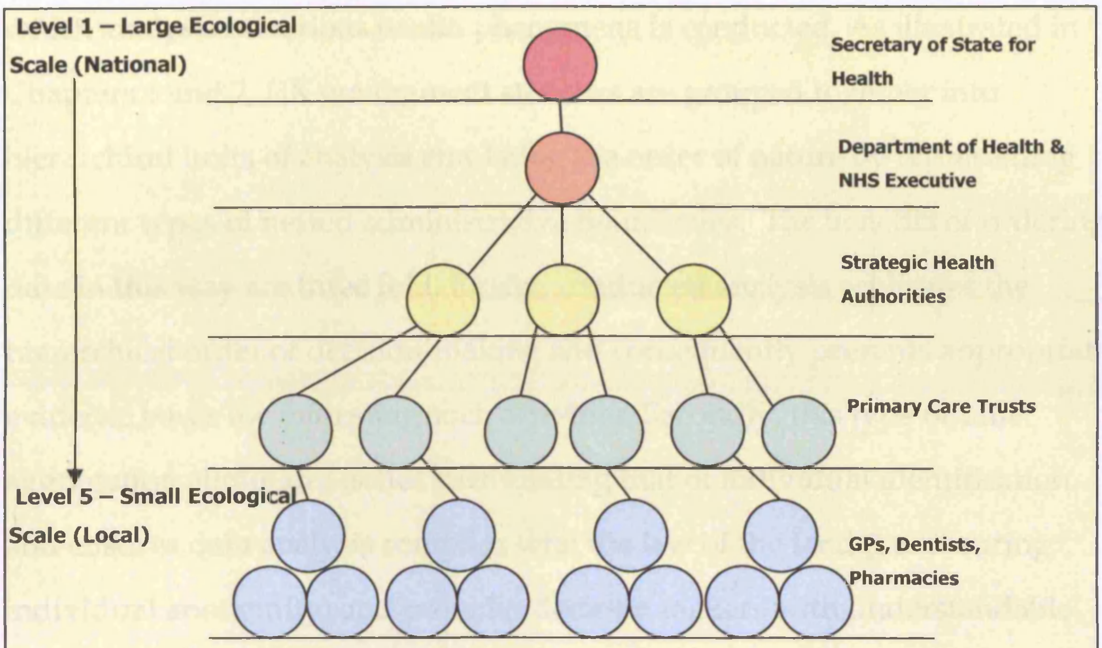


Figure 68: Hierarchical organisation of the NHS

The Strategic Health Authorities (SHA) are the "holons" subordinate to the DoH, who deliver the strategic direction for local areas within their region of responsibility. In 2002 there were 28 SHAs in England, since then a period of rationalisation has seen this number decrease, through amalgamation of existing ones, to 10 larger SHAs (NHS (a), 2007). The responsibility of Primary Care Trusts (PCTs) nests within the SHAs but act individually as a system within a larger system because these organisations are responsible for 80% of the NHS total budget. The lowest level of the hierarchy, level 5, sees the frontline health organisations that deal directly with patients face-to face; doctors, pharmacies and dentists for example. It is also possible to consider the bottom level of the hierarchy, level 5, to be associated with a functional form of scale, as this is where patients are organised according to a functional unit (although this does not correspond to the functional unit of their social space).

It is important to understand the formal geography of the NHS. It is these formal geographical units which predetermine the measurement scales on which analysis of various health phenomena is conducted. As illustrated in Chapters 6 and 7, UK government statistics are grouped together into hierarchical units of analysis emulating the order of nature by representing different types of nested administrative boundaries. The benefits of ordering data in this way are three fold. Firstly, conducted analysis replicates the hierarchical order of decision-making and consequently presents appropriate evidence bases for informing such activities. Secondly, this type of unit aggregation eliminates issues surrounding that of individual identification and ensures data analysis complies with the law of the land (i.e. ensuring individual anonymity) and provides decision makers with understandable statistics once again befitting the structure of their decision-making processes. Furthermore, because these statistics are produced within this ordered data structure, it is feasible to make comparisons across areas within the same hierarchical level. Thirdly, different public sector organisations and data producers have a standard set of boundary data for data production, facilitating and easing the otherwise difficult process of joining disparate datasets.

Thus whilst there are benefits of exploring health outcomes at the formal scale of NHS, invariably there are negative consequences too. The most sizable constraint to this type of spatial aggregation of data into formal geography leads to two widely researched problems in social science and human geography; the ecological fallacy and the modifiable areal unit problem (MAUP). Both of which were referenced in the literature review (Chapter 3), but are deserving of a revisit because the arising complications impact all aggregated statistics of a geographical nature, and are particularly relevant to the nature of analysis in this thesis. The aggregation of data into larger units of analysis creates the classical problem of ecological inference (Weeks, 2004) resulting from aggregation bias.

Observations made at a particular group level are frequently used to make inferences about individuals, but these group level associations are not necessarily applicable at the lower level. This often leads to the development of false conclusions: spatial scale dependent associations are seen at one hierarchical level, but not another. These dependent ecological inferences are commonly referred to as an artifact of the ecological fallacy. Social marketing activities are very much centered upon finding the individual and understanding their behaviour in order to empower, encourage and effect positive behaviour change. Individual fallacies brought about by group level inference would undoubtedly limit the success of social marketing campaigns.

Ensuing analysis for these discrete hierarchical levels returns evident spatial variation with different levels of magnitude, depending upon which formal geography was used for analysis. This issue is known as scale dependent heterogeneity whereby spatial variation at one level may be homogenous but at a different level may be heterogeneous. This issue is compounded by the fact that the boundaries between each level are abstract and fuzzy. Spatial processes are scale dependent, and, as noted by Atkinson and Tate, "the scale of measurement will determine the magnitude of spatial variation", (2000, Page 608). This means any type of geographical analysis, formal and/or functional is simply attempting to model aspects of the real world. Therefore all analysis is just an abstraction of reality.

This reality is consequently predetermined by the methods of data collection and aggregation, chosen methodology and resultant analysis, each of which is primarily determined by the chosen scale of measurement, or the formal organisational geography. Furthermore, because this type of geography implies a sense of unity, the ecological fallacy is manifest during the interpretation of results. False homogeneity is frequently assumed within the areas defined by the formal geography (Moon 1990). Cautious consideration must be attached to the interpretation of any analysis conducted using the formal NHS operational scales, outlined in Figure 68.

The use of formal geographies and their hierarchical scales are not restricted to the NHS data and are evident for most government and public sector organisations in the UK. They produce a workable organisational structure, but their appropriateness for defining resource allocation and identifying and predicting need is limited not only by their susceptibility to fall foul to the ecological fallacy. These formal boundaries are arbitrary. Despite being designed to delimitate boundaries of organisational jurisdiction, they are not prohibitive to patients. Services not provided for within the spatial confines of a particular formal geographical area, will be made accessible to the patient elsewhere. Patients living on the border of these boundaries often have flexibility with respect to accessing services. Patients have a degree of choice when selecting general practices and recent research highlighted their catchments to be geographically diverse and scattered (Jones, 2004). Formal health spaces in reality are fuzzy and complicated to define. As the population becomes more transient, and accessibility to services near place of work not just place of residence becomes increasingly important, these formal geographies will become more and more dissociated from reality. This will render them unhelpful to social marketers because they will not adequately or accurately represent the health phenomena or the health needs of the population. Their ability to provide an abstraction of reality will be significantly reduced.

8.4.1.2 *Local and/or functional geography*

Function scales in geographical inquiry are used to, "delineate breakpoints between spheres of influences in adjacent facilities or features" (Longley et al., 2005, page 135). In this instance, the functional scale attempts to emulate differences in social representations and locate functional areas of social similarity. The functional scale attempts to "maximise interaction within areas and minimise interaction between areas" (Longley et al., 2005, page 135). This research has attempted to homogenise social processes within areas using averages for representing social processes for functional areas of the neighbourhood. This supports Moon (1990) in his consideration of functional geography to be the based around the ideas of the community, neighbourhood or locality and relate to the perceived functions in which the space fulfils (1990, page 166).

The exact definition of what makes a neighbourhood is uncertain. The term has been in use for some time, although no formal definition exists. Local perceptions of neighbourhoods may be defined by natural dividing lines such as roads and rivers, changes in housing design or tenure, or the sense of community generated around centres such as schools, places of worship, shops or transport links. In the past census wards were often used as a proxy for neighbourhoods, not least because statistics were available for them, deprivation measures were created at this scale (see chapter 3, section 3.1). But it is recognised that there are problems because of the large geographical extent of the units and complications of applying group level inferences to the individual (ecological fallacy).

One could consider the catchment of a general practice to represent a type of functional scale representing health outcomes of patients. This was shown in Chapter 7, section 7.4.3.4, even where practices had a predominance of certain neighbourhood Types, their patient lists were by no means socially homogenous. In the literature there has been much discussion about the size and uniformity (or lack of it) of catchments for General Practices. The following reasons were given by a patient for choosing a doctor: Nearness of surgery, desire to maintain link with the doctor the patient knows, woman doctor, cultural group, services offered, specialist clinics. Thus, it follows that the catchments for general practices are often poorly defined (Robson, 1995) and there is geographical overlap. For these reasons it is not appropriate to use catchment of a general practice to define neighbourhoods. But they do present a useful mechanism for facilitating social marketing campaigns that are aspatial. A more detailed review of the conceptualisation of the neighbourhood would be a useful area for further work.

For this research it seemed pertinent that geodemographics provide an alternative representation of the neighbourhood. The use of the neighbourhood, a sociological construct, is more representative of the social space within which people live and interact differently. Geodemographics provide the average social characteristics for a neighbourhood at small ecological scale, ie space as is considered with a particular consumer/patient focus, and space as a resource for social marketing. Although exploration of this functional geography is not without its limitations

The notion of the ecological fallacy and MAUP is not restricted to the formal scale of geographical health inquiry. The very idea that UK postcodes, as used to represent socially homogenous neighbourhoods, which can be classified into 61 population Types representative of the population of one diverse London Borough, is a derivative of the ecological fallacy. Camden has a population composition that is considerably different to 'Middle England'. What is more, some neighbourhood Types (eg Type 40), can be predominantly found in Scotland, and are not representative of England's population. So whilst we saw in Chapter 7 that geodemographic classifications are flexible and adaptable to multiple geographic scales and enable exploration of the functional scale, one must also remember that they are still subject to the ecological fallacy related to social scale because in themselves, the classifications are merely ecological constructs.

All this said, with these limitations being observed, the indicators still facilitate our exploration of social space (from the perspective of individuals as consumers). These representations of social similarity consider space as the medium in which fixed social processes occur (Smith, 2006 on Simmel, 1855) and are reflected in the neighbourhood Types that comprise the classification. But within these neighbourhoods live the patients who create the need and demand for these services. But even with this organisation, the notion of scale is complex. If a patient lives in a neighbourhood, one has to consider where the neighbourhood scale ends and when does it became a borough. Whilst the arbitrary administrative units make this notion of scale more tangible to perceive and understand, the reality is very fuzzy. Is there actually any tangible relationship between the neighbourhood and PCT boundary? This is an epistemological question that could be a subject of further research.

8.5 Combining social and geographical space and scale

The application of hierarchy theory for understanding complexity, in respect of social geographies of health, has been growing in recent years: see, for example, Gatrell (2002, 2004). Indeed, geodemographics classifications created nested hierarchies of population sub-groups to provide an organised view of social space, but are more relevant representations of social processes than formal geographical scales. Working on this view that the functional scale, represented by geodemographic classifications, provides an improved scale for identifying social similarity and similarity in social processes it is possible to consider linking together geographical space with relative social space to explore both geographic and social space in parallel.

Social space was first conceived by Émile Durkheim in the 1890s. During his exploration of social differentiation he was concerned with social morphology (distribution of social forms) and social physiology (interaction and segmentation), (Buttimer, 1969). As it stands today, social space is representative of the scale at which groups live, move and interact, i.e. the neighbourhood. The social construction of scale means there is no characteristic operational scale (McMaster and Sheppard, 2004). Health behaviours of individuals are in part influenced by those around them and those making up their social network (compositional and contextual effects. Neighbourhoods can be used as the central functional unit of analysis and/or scale to assist us to delve more deeply into our understanding of local social context. This is because currently in the UK they are the smallest scale for which data are available and approximately approach social similarity.

From Bourdieu's perspective a number of key constituents interlock together to create social space (Bourdieu, 1997). It is composed of a number of different capitals; economic, cultural and social (see literature review, Chapter 2, Section 2.4.3), which all vary in volume and composition and fluctuate over time. It is the differences between these components that distinguish different social classes and population sub-groups. The differences are derived from, "the overall volume of capital, understood as a set of actually usable resources" (Bourdieu, 1984, page 112). Using this notion of social space and combining it with the information derived from the geodemographic classifications, social space can be explored from Bourdieu's perspective within the realm the health and lifestyle domains using geodemographic classifications and their associated sets of measures.

In 2004 Anthony Gatrell and colleagues took Bourdieu's ideas on social space and social capital and set about modelling the determinants of health inequalities in social space. They began by visualising survey variables in terms of the differential distributions of different forms of capital using multi-correspondence analysis. The types of variables they used as determinants of health inequalities ranged from: gender, age, marital status, educational qualifications , social class to housing tenure, dampness in home, central heating, loneliness and sense of community. The variables were mapped on a two by two grid representing social and economic capital. The study highlighted that whilst people may share geographical space because they are adjacent to each other, their social spaces are dissimilar. This type of analysis highlights a cross sectional view of social space but ignores any temporal aspects (Stafford and Marmot, 2003), it presents a snapshot in time.

Building on Gatrell's work on social and geographical space and Bourdieu's belief that because of the diverse nature of social space distributions must account for, "socially ranked geographical space", (Bourdieu, 1986, pg 124), the final notion explored in this thesis is geodemographics applicability to define a ranked space for both the geographic and social space. This is because as for the determinants of health outcomes the social setting of the individuals and communities are inextricably linked to the geographical settings (Gattrell, 1997).

Building upon the portfolio of indicators created in the previous chapters, together with the variables included in the geodemographic classification, it was possible to create a composite indicator to represent social space for each neighbourhood Type, using proxy variables for different conceptions of capital. The indicator developed is a variant of Bourdieu's notion of different capitals, called here 'lifestyle' status. It is calculated by averaging together a number of indicators. Economic capital is represented by unemployment, salary and housing tenure. Cultural capital is represented by level of education attainment whilst the social capital, known as bonding capital, is represented by the different classifications of neighbourhood Types. The result of averaging together these scores is a composite (equally weighted) indicator of lifestyle status for each neighbourhood Type. Further work would include creating a more intelligent method of composite indicator. A low score symbolised a low lifestyle status for the corresponding neighbourhood Type.

In a previous section of this chapter it was observed that the indicators developed during the analysis phases of the thesis were most appropriate for diseases relating closely to lifestyle. For this reason the composite indicator for health status was created by averaging together the scores for conditions and diseases which had a weighted deviation score (section 8.2) greater than 50. This was because it represented health outcomes that can actually be discriminated by the neighbourhood Type and could be linked both to individual and group behaviours. A high score characterised a neighbourhood Type with a high health status.

The two status domains, lifestyle and health, were then plotted on a graph for each neighbourhood Type. The populations of Camden residing in each Type were also plotted, with larger circles representing neighbourhoods with greater proportions of the population. The result is illustrated in Figure 69. This graph provides a visualisation of Camden's population within their social space. The y-axis represents the health status of each neighbourhood Type, and the x-axis the lifestyle status. The colour of each graduated circle equates to the nested hierarchical neighbourhood Groups that organise the Types according to ecological hierarchy principles previously discussed.

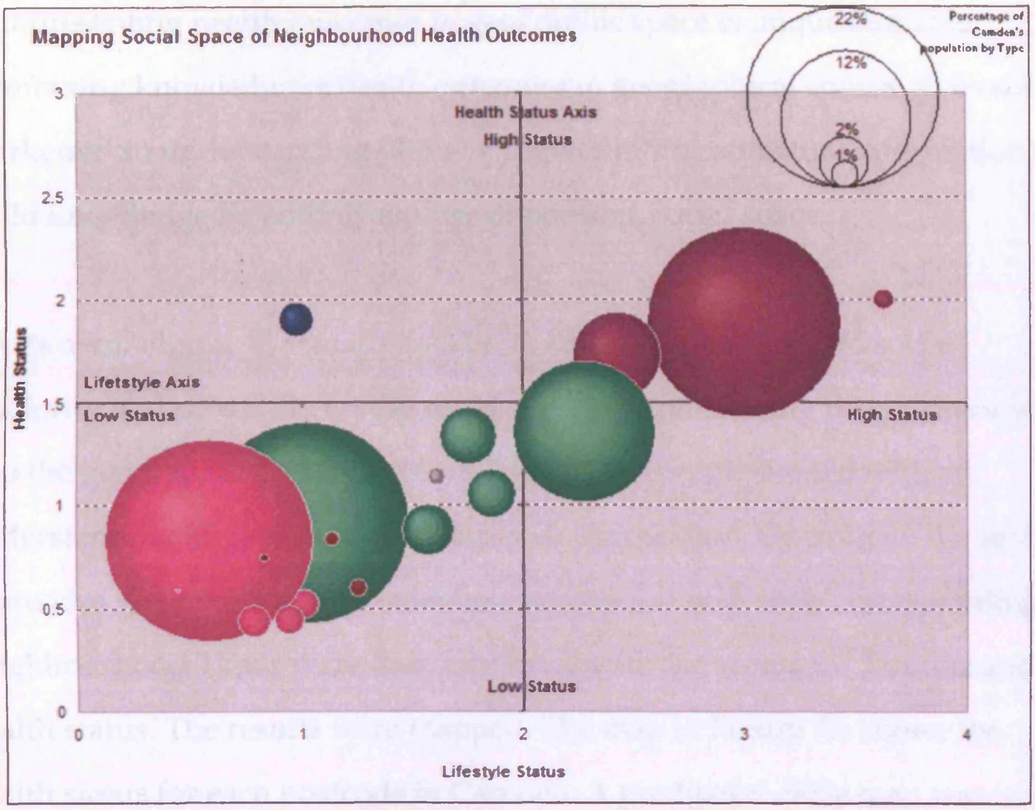


Figure 69: Camden neighbourhoods plotted in social space

Figure 69, highlights an alternative visualisation of the composition of Camden's population. By and large, neighbourhood Types that belong to the same hierarchical groups appear to have similar lifestyle and health statuses. No neighbourhood Types fall into the bottom right quadrant of "high" lifestyle status combined with "low" health status. In general the gradient of inequalities proposed by Marmot and Wilkinson (2004) and discussed in the literature review (Chapter 2, Section 2.3), is apparent. One neighbourhood Type is not located on the gradient. This is Type 08, representing new housing and families just moving in. These young families have a low lifestyle status but rank highly in the health domain, they are often found in new builds and therefore there is often an absence of geodemographic data for such Types.

Figure 69, visually represents health status in social space for neighbourhood Types, but it does not represent either in geographical space. The importance of representing health outcomes in geographic space is unquestionable. Combining knowledge of health outcomes in geographical spaces gives social marketers an understanding of place. It gives more contextual information to build knowledge by adding another dimension, social space.

On its own, Figure 69 visually transmits information relating to social similarity and proximity on the social scale. By transferring this information into the geographic domain, knowledge can by complemented with an understanding of geographic similarity or polarisation. Once again the unit postcodes were used to approximate a neighbourhood, their corresponding neighbourhood Types were then attributed with the scores for lifestyle and health status. The results were mapped. The map in Figure 70, shows the health status for each postcode in Camden. A graduated circle map was used to map the two domains; lifestyle and health status. The size of the circle represents the health status of a neighbourhood, the larger the circle the higher the health status. The colour of the circle indicates lifestyle status: a graduated colour scheme from yellow to red equates to the movement from a "low" lifestyle status to a "high" lifestyle status.

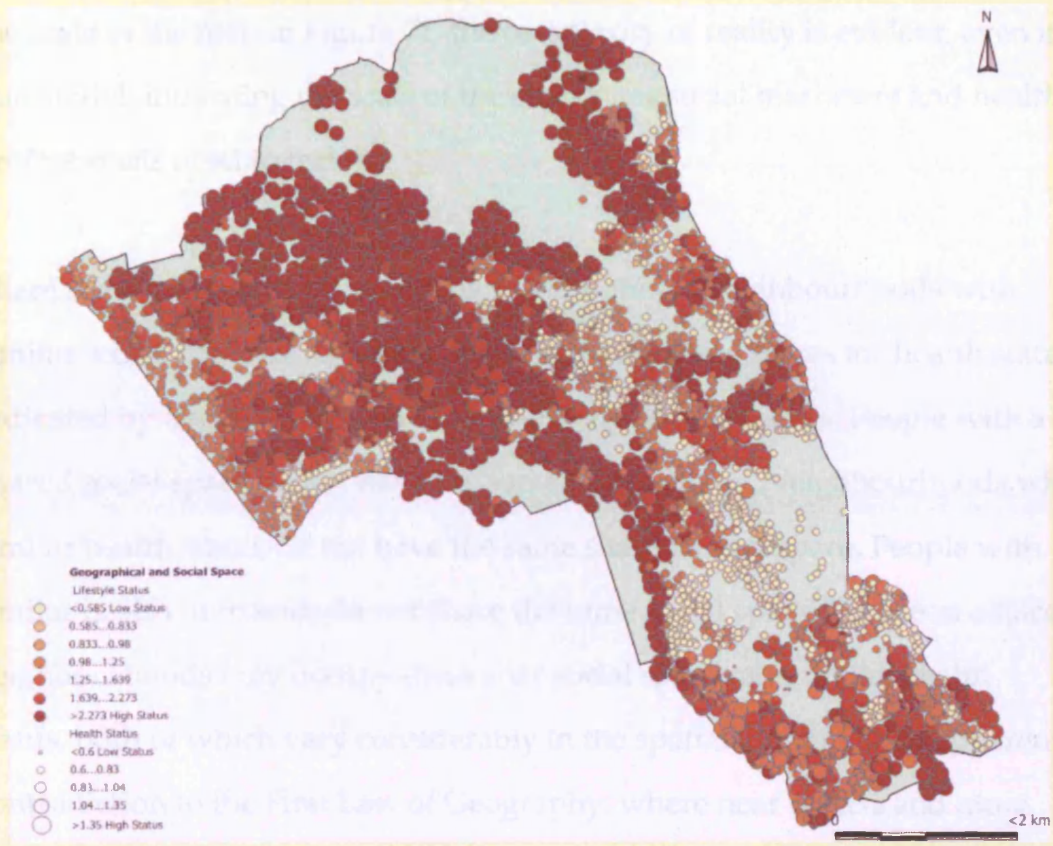


Figure 70: Representations of geographical and social space for Camden's neighbourhoods

One of the striking features of the map is the considerable diversity and polarity that exists in Camden's social space with widely differing health status of the population. The maps visually capture the inequalities. Large disparities in social space and health inequalities are evident for nearest neighbours – does this contradict Tobler's First Law of Geography? It is certainly evident from this manifestation of social and geographical space that nearest neighbours are sometimes vastly dissimilar. Figure 71, zooms into an area in the centre of Camden. The smaller spatial extent of the map, allows a clearer and more detailed examination of the vicinity and its composite neighbourhoods. The overview map in the bottom right hand corner indicates the location of the smaller geographically scaled map. Once again the same coloured circles represent neighbourhoods that are expected to have similar lifestyle status: those with a high lifestyle status are in red.

The size of the circles denotes the expected health status from low to high. At the scale of the map in Figure 71, the complexity of reality is evident, even in this model, indicating the scale of the challenges social marketers and health professionals need to tackle.

There are a number of observations worth noting. Neighbourhoods with similar scores for lifestyle status do not have the same scores for health status, indicated by circles of the same colour having different sizes. People with a shared social space do not have the same health status. Neighbourhoods with similar health status do not have the same shared social space. People with similar health outcomes do not share the same social space. People in adjacent neighbourhoods may occupy dissimilar social spaces and unlike health status, both of which vary considerably in the spatial domain. This apparent contradiction to the First Law of Geography, where near objects and more related than far objects, could this have come about because of the housing and occupational policies of the local council?

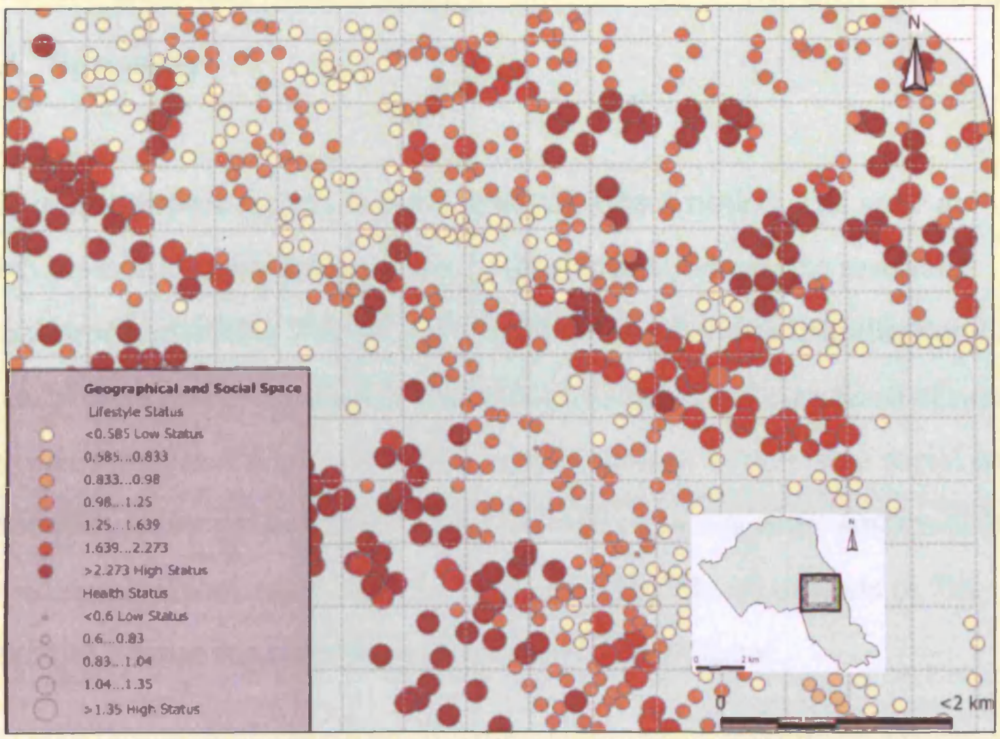


Figure 71: High resolution map of geographical and social space combined with health status in Camden

The results of combining analysis of social and health status in the geographic domain illustrates the impact of scale on the complexity of understanding health and social status. The higher the geographic resolution used to explore geographical scale, intuitively the more complex the model becomes. The measurement of the scale and extent of the differences seen in the visualisations are not in the scope of the thesis, but conducting a LISA analysis of the neighbourhood points would be a useful exploratory tool for further work. It would provide social marketers with knowledge of concentrations of similar neighbourhoods based on health status and social space. It would also indicate isolated neighbourhoods that are most dissimilar to their surrounding neighbourhoods. From a social isolation perspective, this knowledge is important for health professionals and enables them to acquire a detailed picture of local knowledge, in the absence of direct personal experience.

8.5.1 Summary

Scale in geographic inquiry is not a straightforward notion. The scale at which phenomena are measured and conceived influences the resultant spatial representations. This section explored how the conceptualisation of space, be it formal or functional, will produce different abstractions of reality. The use of the functional scale of the neighbourhood to represent social space and socially homogenous populations and their ensuing representations has proved a useful tool, extending the initial indicator developments of this thesis into a more conceptual representation of local space.

If the different types of capital identified by Bourdieu to encapsulate social space are considered as a resource from an economic perspective, they could be measured as a stock, which is what this research does. These stocks of economic, social and cultural capital with inputs and outputs comprise the social space and levels of these stocks are influenced by the complexities of reality in which the social space is located in geographical (and temporal terms). By concentrating on defining a lifestyle domain comprising of these stocks and identifying a stock called health status, it was possible to locate Camden's neighbourhoods within a multi-dimensional space.

This enabled the representation of the uniqueness of different places in Camden as more than artificial geographic or sociological constructs. Providing useful visualisations of the real context for communication, exchange and decision-making required for the social marketing process.

8.6 Chapter summary

In this chapter, some of the outcomes derived from the analysis in the case studies of Chapters 6 and 7 were considered in greater detail. The case studies produced a series of interesting results that were worthy of greater scrutiny. The studies presented in the previous chapters built upon and substantiated a geo-spatial and geodemographics framework using a series of policy relevant applications useful for social marketing campaigns. This chapter has provided a more detailed discussion on a number of the interesting observations derived from the main outcomes of Chapters 6 and 7.

The application of the framework to the case studies has appeared to open up a series of new dimensions that were discussed in this chapter and has identified potential opportunities for further work across a number of these themes. Firstly, it has considered the notion that the geodemographic classifications provide a robust representation of the population distribution of sub-groups in England. This section considered the homogeneity of social class within the different Types as indicated by responses calculated from the Health Survey for England. The Simpson's index of dissimilarity was used to calculate how dissimilar counts of social class were within neighbourhood Types. Only one neighbourhood had a useful score of dissimilarity. This highlighted the importance of further work that is required to understand the effects of the averaging processes used to create index scores themselves. It signals a requirement for a more detailed inquiry into the accuracy of the social homogeneity that these classifications are considered to represent. One limitation that arises from this averaging process is a reincarnation of the ecological fallacy because the classification represents an ecological scale. So, whilst geodemographics used at their lowest geographical scale, the neighbourhood Type projected for unit postcodes, removes issues emanating from the modifiable areal unit problem they are still subject to the ecological fallacy and are thus not infallible.

Whilst being aware of the limitations and the need for further work in this area, the second chapter considered the effectiveness of the geodemographic measures to efficiently distinguish the health needs of contrasting neighbourhoods. Total weighted deviations were calculated for each of the created measures to determine how much each neighbourhood actually deviated from the national average and to calculate a total score for each individual disease of lifestyle condition measure. The results confirmed that geodemographic classifications are useful at differentiating need relating to health-harming behaviours but are less useful for diseases where determinants are less strongly associated to lifestyle and behaviour choices. Therefore, with respect to social marketing campaigns they have proved themselves a useful quantitative tool for locating population need.

The third section of the chapter addressed an issue that was first identified in the literature review (Chapter 3), noting the limitations of traditional deprivation measures, first as merely population containers and second because they are not specific enough. Many conditions are linked to a gradient in health outcomes and the rankings of arbitrary formal organisational areas based on a composite measure is only successful at highlighting the very best and the very worst areas, the ranks of areas in the inter-quartile ranges are not conclusive. The geodemographic framework presented in Chapters 6 and 7, illustrated their usefulness for exploring health inequalities from a different perspective. The penultimate section of this chapter took the health deprivation domain indicator from the government produced Index of Multiple Deprivation (2004) and applied it to the geodemographic classification to explore how useful they were at distinguishing health deprivation for neighbourhoods. The results indicated that diseases or lifestyle conditions could be appropriately targeted using the health deprivation scores from the IMD 2004 for teenage pregnancy, asthma, influenza and smoking. So for this purpose, traditional deprivation measures have their place in research but alternative approaches can, as suggested by this thesis, present a more enriched picture of health inequalities ratifying their appropriateness for use in social marketing.

The final section in this chapter discussed the conceptualisation of different scales of geographic inquiry that were used throughout the case studies. The way in which phenomena are measured and conceived naturally influences the resultant spatial representations. The measures created in Chapter 7, illustrated this point, showing that the measures produced at different scales returned different abstractions of reality. In the final section, the functional scale of the neighbourhood was used to represent social space and its socially homogenous populations. The ensuing cartographic representations proved a useful tool, extending the initial indicators developed in this thesis into a more conceptual representation of local space which was they applied to understand the health status of residents in Camden. This indicated that the functional scale of the neighbourhood was a more congruent scale for observing the relationship between population sub-groups, their interaction and their actions and behaviours.

Section III

Chapter 9

Results dissemination and knowledge transfer

9 Results dissemination

9.1 Introduction

This chapter marks the beginning of the final section of the thesis, deliberating how the research can be turned from research knowledge to professional knowledge. The previous part, section II, was concerned with the development of an alternative geodemographic framework for exploring health inequalities and distributions of health outcomes from the perspective of a social marketer. In section II a number of commonly used health datasets were incorporated into the framework and subsequent empirical analysis conducted to understand its real world application for social marketing. During the research alternative health views of Camden's local population were developed, representing health outcomes for the functional geographical scale of the neighbourhood. It was evident that the framework has potential to be very useful to health professionals.

The research and discussion that is documented in this chapter realises the third and final objective of the thesis which was to:

> **Embed and evaluate framework in public health thereby transferring research knowledge to professional knowledge**

On reflection, this objective is concerned with the evaluation of the indicators with respect to how they can be embedded into existing health promotion and prevention initiatives in the health professional world. The framework and measures developed during the course of completing objectives one and two should be disseminated and best practice shared with public health practitioners. Objective three sets out to explore and implement a set of tools for communicating and distributing appropriate information to different audiences.

One key output of all the research conducted in section II was the development of a series of cartographic visualisations, representing a range of complex health phenomena. For these visualisations to be disseminated and shared the next step in the research was to explore methods of cartographic communication for public health and social marketing purposes. To maximize the communication potential of the maps this chapter explores mechanisms for distributing the tools and techniques, with relevance to public sector departments and organisations. This chapter considers the general uptake of geographical information and their relevant systems (GIS) in the NHS, and in light of available expertise and experience it explores the dissemination of the geodemographics framework and its associated visualisations within the public health department of Camden PCT.

Section II of this thesis highlighted a wide range of explorative and practical applications of GIS and geodemographic technologies that could be practically applied to aid social marketing intervention planning. To maximise their use and raise awareness of such innovations within the NHS, this next chapter delves into the practicalities of GIS technology diffusion, implementation and communication within the NHS. The chapter outlines the diffusion process adopted for disseminating this research and explores both its limitations and successes. Once again, Camden PCT was used as the test-bed for this technology to roll out a simple web-based repository for cartographic visualisations of a selection of the measures developed during the empirical investigations of the previous section.

A number of strategies were adopted to encourage the practical uptake of ideas and encourage long term use of the techniques. The first strategy comprised of a series of seminars aimed at increasing awareness and knowledge sharing. The second was to produce a variety of tangible products that could be easily used by practitioners for decision making; a printed health atlas of Camden was developed alongside a web-based interactive tool. This chapter discusses the techniques used to raise awareness of the project and diffuse GIS and geodemographic technologies into the NHS environment.

Before exploring the practicalities of GIS implementation in the context of Camden PCT, it is first relevant to discuss the notion of GIS technology diffusion/implementation in the NHS. It is beyond the scope of this thesis to review the entire literature on this topic but key papers are worthy of consideration. This is because they assist in contextualising both difficulties and limitations associated with implementing new technology innovations in large, highly controlled hierarchical organisational structures such as the NHS.

The common perception of the term technology is in line with one of its definitions in the Oxford English Dictionary that it is, "machinery and equipment based on scientific knowledge". The dictionary also acknowledges that the term corresponds to "the application of scientific knowledge for practical purposes". Thus, both in the technical definition and practical usage of the term there are two equally valid meanings.

Masser and Campbell (1995, page 6) reiterate the ambiguity which surrounds the usage of the term technology. They note that it is comprised of three core components: machines, methods and knowledge. The definition of technology adopted in this research encompasses not only the machinery and equipment but the scientific knowledge and its application. It can be considered as: desktop personal computers (PC), data, GIS and geodemographic methods and research knowledge relating to these methods, key components outlined in the research methodology.

On their own, PCs and GIS/Geodemographic methods may not be considered as innovative components. As explored in Chapter 3, geodemographics technologies have been used in commercial practice for a number of decades, but the process of combining their use in conjunction with operational health data, surveys and individual knowledge creates an innovative process which is one of the unique features of this PhD. Facilitating new capabilities within a public health setting and engaging end users in alternative and creative techniques for exploring rationed based national health care.

Technology diffusion is commonly studied "as a process by which older, outdated technologies are replaced by more advanced, efficient and hence more beneficial ways of doing things", (Wegner and Masser, 1996, page 9). The diffusion of geodemographic GIS technology within a health care setting can be considered as more advanced and efficient because of the enriched picture of the population it presents. Although they still require further validation of the true level of their accuracy, they do facilitate an approximate representation of social space of the neighbourhood. Their use in the health care setting can potentially assist health professionals to identify specific sub-population groups and their corresponding health needs. This encourages the implementation of more appropriate localised interventions and social marketing campaigns that are representative of local people. In the past, existing public health policy and practice was concerned with broad and universal implementations, providing equal services for all, an approach now considered to be both too costly and inefficient, hence the notion of ration-based healthcare.

The literature review, chapter 3, noted the suitability of a GIS and geodemographic framework to perform a variety of functions, used to inform evidence based decision making within the UK health sector and resurgence in their use within the public sector. GIS lends itself to the role of analysing spatial patterns of disease, uptake of services, understanding health inequalities, developing social marketing campaigns, needs assessment and service accessibility to name but a few, see (chapter 5) . With the onset of web-based internet tools for disseminating information and data, these applications of GIS should become more accessible to professionals and the public alike. Higgs et al (2005, page 105) note, "the increasing availability of web-based tools that enable the integration of spatial data ...also encourage public access and input into public health decision making as well as providing professionals with real- time information".

Building upon their past research into the area of GIS and its potential for the NHS (Higgs et al., 2001, 2003), Higgs et al (2005) explored the use of GIS in the UK NHS. In 2001 questionnaires were sent to all information technology/services personnel working across various health authorities and boards across the United Kingdom. The results of this were analysed and a sub-set of respondents were then questioned using a semi-structured interview. A number of human, cultural and organisational constraints were identified as barriers to moving GIS beyond simple map production.

The biggest hurdles identified by the authors corresponded to work time constraints, insufficient staff and lack of budgets, lack of training, GIS failing to meet analytical requirements, lack of digital data and a lack of demand to increase GIS use. My personal observations witnessed a number of these barriers present at Camden PCT over the period 2004 to 2007, during the implementation of the research project. In the absence of a geographic information strategy, the power of GIS was lost; it was seen merely as a, "function to produce pretty maps" (personal communication: health promotions strategist: July 2006).

Despite the fact that 85% of the health authorities surveyed by Higgs et al 2001 were using GIS in some form of another, indications from these results and from personal observations at Camden PCT (through 2004 to 2007) suggest that the practical application of GIS is severely limited within NHS organisations. In 2001 Higgs et al noted the divide between applications of health GIS in academia and the practical mainstreaming of these ideas into the NHS. A number of explanatory factors were presented to explain this divide.

GIS organisation strategies are absent from both national and local NHS plans. Its use remains uncoordinated and is rarely used to inform strategic planning. Boulos (2004, page 43) calls for the NHS to, "...start by carefully defining the purpose(s) of a nation-wide, coherent GIS implementation across its organisations, and by developing a clear "GIS business plan". Present and past calls for greater utilisation and diffusion of GIS technologies within the NHS remain, having had little impact on its diffusion as a whole. The adoption of GIS software has occurred, in terms of purchasing GIS software, but moving from acquisition to utilisation beyond the realms of map production has not advanced. So whilst there remains significant potential for GIS within the NHS, realising this potential has been and remains problematic.

Over the last 15 years a considerable array of research has presented an evidence-base to support the application of GIS technologies within the public sector. For example, the second volume of the book Geographical Information Systems: Principles, Techniques, Management and Applications, referred to as, "the big book of GIS" by Longley et al (1999) presented a range of real applications for health, police, education and local government. But despite the overwhelming evidence to support GIS diffusion, even in the broader context of public policy and local government it still remains notably absent from government white papers and strategy. Its absence has not passed without recognition. Higgs et al (2003) remarked upon the government white paper of 1999, which laid out the agenda of 'joined-up' local partnerships, made no mention of the role of spatial information and GIS. Without local or national policy, GIS technology diffusion and implementation and its skills within the NHS is likely to stagnate, as in past years this is what seems to have occurred. To overcome the lack of policy and to ensure successful implementation it becomes relevant to first understand the epistemology of technology diffusion and implementation, to determine the most suitable methods for diffusion in complex work environments such as that presented by the NHS.

Masser and Campbell (1995) note the complementary nature of diffusion and implementation. Diffusion in the sense of technology diffusion is concerned with the dissemination of technical information and knowledge leading to the adoption and utilisation of new technologies by users. Complementary to this is notion of implementation which is "the process for transforming unproven potential into a taken-for-granted component of daily activities of an organisation" (Masser and Campbell 1995, page 25). These ideas imply that successful GIS technology diffusion requires consideration not only for the advancement of machines, methods and knowledge but for interaction between the technology and the context within which it is located (Campbell, 1996, page 23), this collective concept is referred to as "social interactionism", is discussed in the following sections.

9.2 Results dissemination

9.2.1 Situational context of Camden PCT

Within the GIS technology diffusion discourse, Campbell and Masser (1995) identify two key organisational considerations that ease the implementation of GIS technology. Organisations should consider how best to integrate a new working practice into existing cultural traditions and norms and understand how subsequent generated information will contribute to the decision-making process. Consideration of these factors should assist successful diffusion of new technologies. They note technology implementation is frequently managed by one of three processes: technological determinism, managerial rationalisation or social interactionism.

Technology determinism believes the "inherent technical worth of a particular innovation .... and the process, adoption and utilisation of an innovation is guided by the inherent value of the technological innovation" (Campbell and Masser, 1995, page 27). Implementations that were carried out using technology determinism and that used the technology determinism framework, rely upon the added value of the technology to provide momentum for its implementation. Yet Higgs et al (2005, 2001) noted improved advancements in machines, GIS methods and data (at the time of writing an Ordnance Survey (OS) pilot scheme across the NHS is in place to provide NHS Trusts access to OS data) has had limited impact on the advancement of GIS technology in the NHS. Advancements in GIS technologies (seen here as technology determinism) alone are not suitable drivers for stimulating innovation in the NHS. For this reason technology determinism was not a suitable implementation framework for disseminating the geodemographic and GIS related technologies developed in this research.

An alternative implementation framework known as managerial rationalisation was also disregarded. Implementations using this type of framework refer to innovations where technology potential is recognized as a consequence of strategic vision and planning. Thus, corporate strategy acts as a motivator to provide the impetus for implementing new technologies. Once again, using the NHS as an example, spatial strategies in policy and strategic frameworks are noticeable by their absence. National and local planning frameworks are still devoid of any spatial strategies (Higgs et al., 2003 and Boulos, 2004). As observed at Camden PCT, local strategic plans did not prioritise or even consider GIS within the decision making process. This top-down management approach was unsuitable for Camden PCT because of a lack of any spatial strategy that should be in place before thinking about GIS.

Issues associated with the implementation of new technology innovations in the NHS are not only restricted to GIS. Under the guise of the management rationalisation process of implementation, the NHS IT upgrade which is predicted to cost 12.4 billion pounds has been fraught with difficulties. As a result the project was subjected to government investigation. The release of the Public Accounts Committee report in April 2007 was most damning in its high-level assessment of the National Programme for IT (NPfIT) for many different reasons including, inhibiting innovation and progress and overall lack of significant clinical benefits.

The evidence of previous research demonstrates the limitations of technological determinism and managerial rationalisation as a potential implementation framework for propagating this research in the NHS. These frameworks are inappropriate for spatial implementations within the public health department at Camden PCT at present because of, the lack of a local spatial strategy, distrust and unawareness of new technologies and the techno-phobic nature of many staff. For these reasons the implementation framework known as social interactionism was adopted for disseminating the results of this thesis.

The social interactionist implementation framework places importance on exploring technology within the confines of the society within which it is being implemented. In this framework attempts are made to understand the complex nature of social interactions that are associated with information technology (and GIS technologies) and explore the potential barriers and pitfalls to the implementation process from both a social perspective and a technological view point.

This viewpoint of the implementation process recognises the importance of organisational environments as contexts, and the need to ensure implementations are not treated independently of these situation contexts. This particular framework places emphasis on understanding the situational social context within which the technology is being implemented. The framework was chosen because of its philosophy of user-centred design, sensitivity to change and awareness of cultural uniqueness.

To understand the situational context of a social interaction implementation framework, a number of key organisational and social considerations have to be explored in the situational setting of the technology implementation. The contextual elements to be evaluated comprise of existing user knowledge and technical constraints. Bearing these considerations in mind the next section briefly appraises the situational social context for which this research is being embedded at Camden PCT.

9.2.2 Technological considerations

9.2.2.1 User knowledge

At Camden PCT the responsibility for GIS and spatial data sat within the Public Health intelligence team, at full capacity the team had 6 full time members of staff. This included one dedicated GIS analyst, qualified at postgraduate level. During the duration of this research, between September 2004 and April 2007, the team membership fell to 3 staff, none of whom had any professional GIS skills or qualifications and of which only 1 team member had self-taught skills to carry out basic map production functions using MapInfo GIS.

The team/public health department had no health intelligence, information or GIS strategy. Any higher level information technology strategies present at the PCT were concerned with developing/integrating systems for monitoring GP performance. GIS implementation difficulties were exacerbated by lack of understanding of organisational priorities and how technology contributes to their achievement (Campbell and Masser, 1995). Therefore implementation of GIS and geodemographic technologies within this culture were problematic; what is more, this thesis set to rollout these new technologies not simply for use by the health intelligence team but by the public health department as a whole and interested parties across the PCT. GIS knowledge beyond the health intelligence team was restricted to using maps for inclusion in reports, and could thus be classed as an absent capability. This is because the organisation and the people in it were not used to spatial thinking and were unfamiliar with the practical power of the technology or how to incorporate it into their day to day thinking.

In general across the public health department, advanced computing skills were lacking by the vast majority of staff. PC use was predominated by tasks relating to emailing, report writing or compiling spreadsheets with the occasional use of pivot tables.

From a users perspective the limited use of GIS technologies and the minimal use of PCs in general ruled out the possibility of customising MapInfo to produce a bespoke interactive social marketing tool, as there would be no possibility for skills transference for mainstreaming the product after this research was completed. There was a clear requirement to develop a user friendly application with minimal technical know-how expected from the end-user. In this instance the end-user was considered to be both health intelligence staff and public health practioners across the PCT.

9.2.2.2 Software (GIS)

In 2002/03 Camden PCT purchased a GIS package known as Health-Pro, a product developed and marketed by MapInfo. It combined MapInfo's GIS software with some Ordnance Survey datasets: postcodes (Codepoint) and street network (Meridian). Its utilisation was restricted to simple mapping of census and health datasets. Maintenance contracts for the software or the data were not included in the purchase or any subsequent departmental budgets. Thus, software and data capabilities deteriorated. The lack of budget also inhibited decisions to use MapInfo to develop an interactive social marketing tool because no further licences of either MapInfo or the geodemographic data could be purchased to make the tool available to other members of the health promotion department. As a researcher based within UCL, Mapinfo version 8.5 was used, but the PCT was still utilising an old and now unsupported version 6 to carry out general map production.

9.2.2.3 *Hardware*

Hardware infrastructure including server and data storage at the PCT was restricted. One server was used by all members of staff throughout the public health department. Capacity planning of disks was poorly managed, numerous times the server's storage limit was exceeded, resulting in staff being requested to delete and move documents and data. Thus, options of implementing a central data repository for GIS data, whilst preferable, was not logistically possible due to lack of server capacity especially in view of the GIS file size and outsourced IT department which would only install servers with Microsoft products.

The web server that held the Camden PCT intranet pages was accessible only via a content management system, which enabled development of web pages using a light version of Microsoft Word or by editing HTML code directly. The IT support team would not install/support any non-Microsoft software onto the servers, thus eliminating the possibility of installing MapInfo's web-GIS discovery server, which had been purchased as part of the Health-Pro package in 2002/03, which to this date had remained unused in its original box. A web-based interactive GIS service accessible to all in the PCT was considered to be an unviable option. Therefore the only option available which would enable the collective and systematic dissemination of cartographic information and representations would be via an intranet site. This site could be developed using the maps presented as image files and uploaded onto a general all purpose intranet server.

9.2.2.4 *Data availability*

In November 2005, the Ordnance Survey working with the Information Centre for Health and Social Care launched a pilot agreement to widen access and support NHS adoption and utilisation of geographical data (Ordnance Survey, 2006). Various Ordnance Survey products were made freely available, with no direct costs for PCTs and a range of NHS organisations. The pilot was expected to run until spring 2007. Despite Camden signing up to this agreement, GIS utilisation still remained at simple thematic map production, primarily due to resource constraints and unawareness of its associated capabilities.

9.1 Web-based mapping tool for social marketing

The technological considerations outlined above were a useful starting point to define user requirements for developing a web-based mapping tool for public health staff. By using the social interactionist framework to contextualise the public health world, it was possible to understand more clearly the contextual working environment at Camden PCT.

9.2.3 User requirements

The key users of the system were to be members of staff who work within the public health department, in particular health promotion staff. Health promoters in Camden PCT plan, implement and evaluate policies and strategies to promote health within a specialist setting. Often they work in respect to a specific issue, or within a particular population sub-group. They need information to assist them in understanding the health needs of the population. The mapping tool, derived from the empirical geodemographic framework and analysis will provide health promoters with the following:

- Assist health promoters to gain an enriched picture of the demographic groups living in Camden, providing information on different age groups, ethnicity and social and economic characteristics;
- Provide health promoters with details of the distribution of health needs for neighbourhood and general practices across the borough, for a number of different diseases and conditions;
- Provide health promoters with information about local community facilities;
- Deliver an enriched picture of local neighbourhood health need beyond standard deprivation measures;
- Provide a framework for identifying local neighbourhoods at risk; a means of highlighting neighbourhood variation in likely health outcomes;
- Deliver an alternative tool for understanding service equity (services reaching those with the greatest need);
- Deliver an intelligence led solution for social marketing campaigns;
- Provide a means of describing the health trajectories of small localities to explore how lifestyle may be represented using geodemographic indicators.

9.2.4 Functional and technical requirements

A meeting took place with the PCT communication department in May 2006 to establish a protocol for creating a website for disseminating geodemographic information relating to Camden PCT. As noted, the PCT use a content management system to automatically create a series of hierarchically linked pages, using a parent and child relationship between web-pages. In order to ensure simplicity and usability, the web pages for the Camden Profile utilised the content management system to create a structured hierarchy of pages.

The page structure used is shown in Figure 72. The adoption of content management system to implement the profile reduced the need for extensive technical requirements, because the simplest solution was chosen to fit in with the situational context.

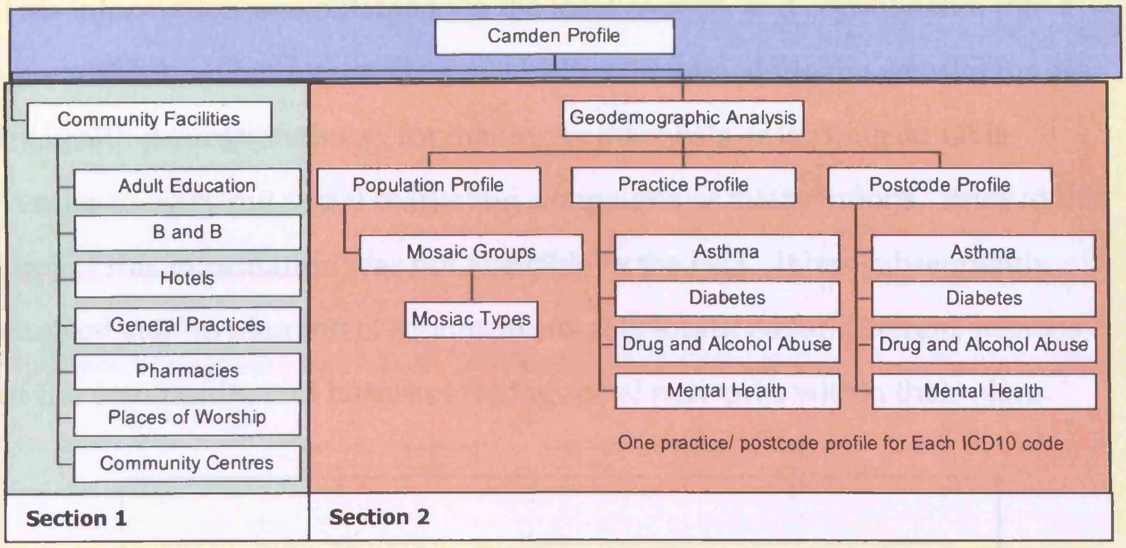


Figure 72: Structural layout of the Camden Geodemographic Profiler

9.2.5 Website design and implementation

The website was accessible through the public health pages on the NHS intranet. The parent directory, "Camden Profile" provided links to the two main sections.

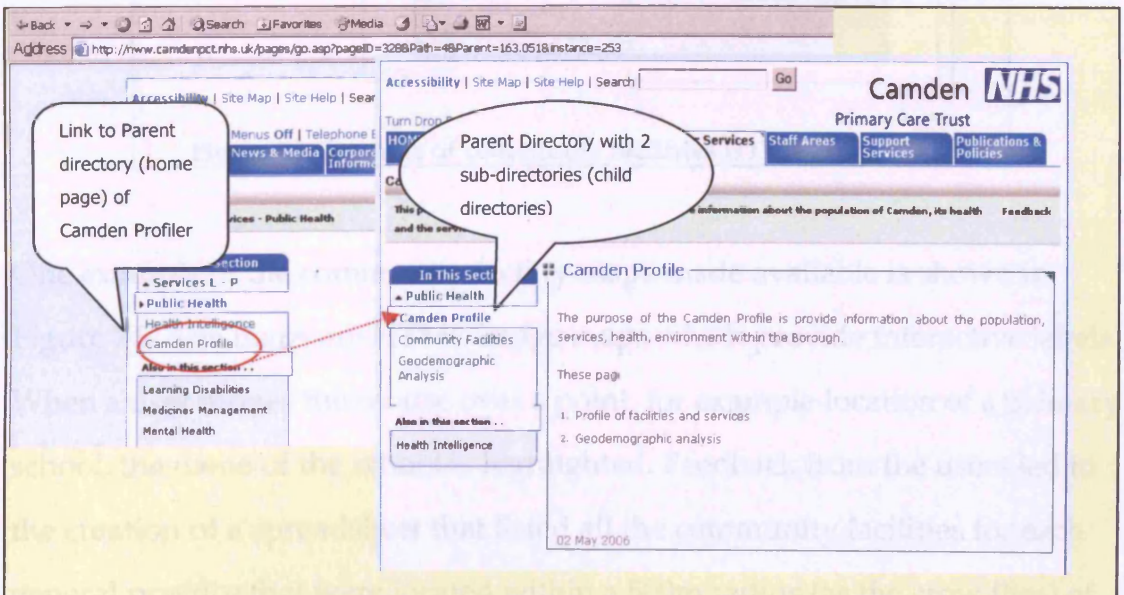


Figure 73: Home page of Camden Profile on NHS intranet

Within the Camden profile, the website was designed to have two key sections (as show in Figure 73). The first section provided staff with information about the spatial location of community facilities in Camden. This information was obtained via the local council, and consolidated into a spreadsheet, which was made available for download via the website for use by health professionals e.g. for mailing/contact lists or locating suitable venues to carry out social marketing campaigns or interventions. Prior to this project this information was not available to the PCT. It has subsequently enabled staff to plan interventions in suitable locations for different sections of the community and harness existing social networks within their plans.

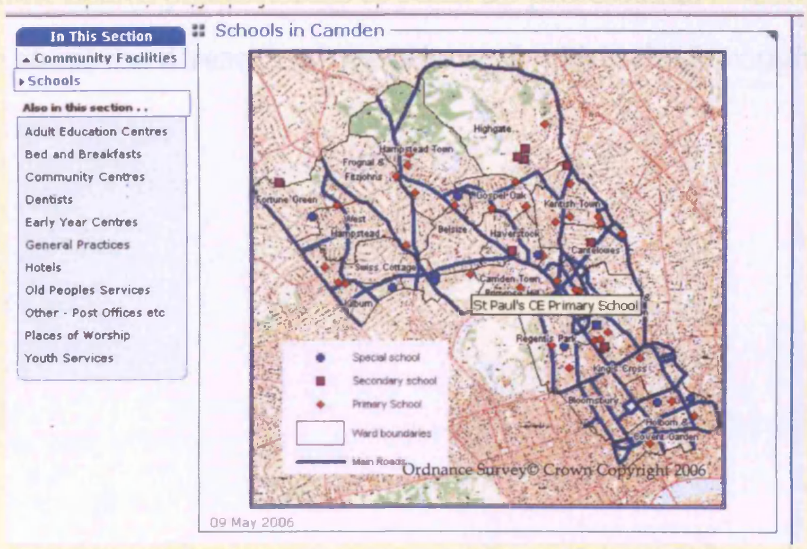


Figure 74: Example of community facilities HTML image map

One example of the community facility maps made available is shown in Figure 74. The maps are HTML image maps which provide interactive labels. When a user moves the mouse over a point, for example location of a primary school, the name of the school is highlighted. Feedback from the users led to the creation of a spreadsheet that listed all the community facilities for each general practice that were located within a 500m radius (as the crow flies) of the spatial location of the practice. Once again this was which was made available for download in spreadsheet form to be used for intervention planning.

The second section of the website presented specific geodemographic information relating to the local population of Camden. It was divided into three sub-sections; population profile, practice profile and postcode profile, as highlighted Figure 75. The population profile contains information about Camden's neighbourhood populations, based on the framework developed in Chapter 7. The geodemographic profile provides information pertaining to the average characteristics of each of the different neighbourhood Type represented. The rich body of information contained within these pages, enabled public health professionals to build up professional research knowledge about the diversity of needs located within the Borough.

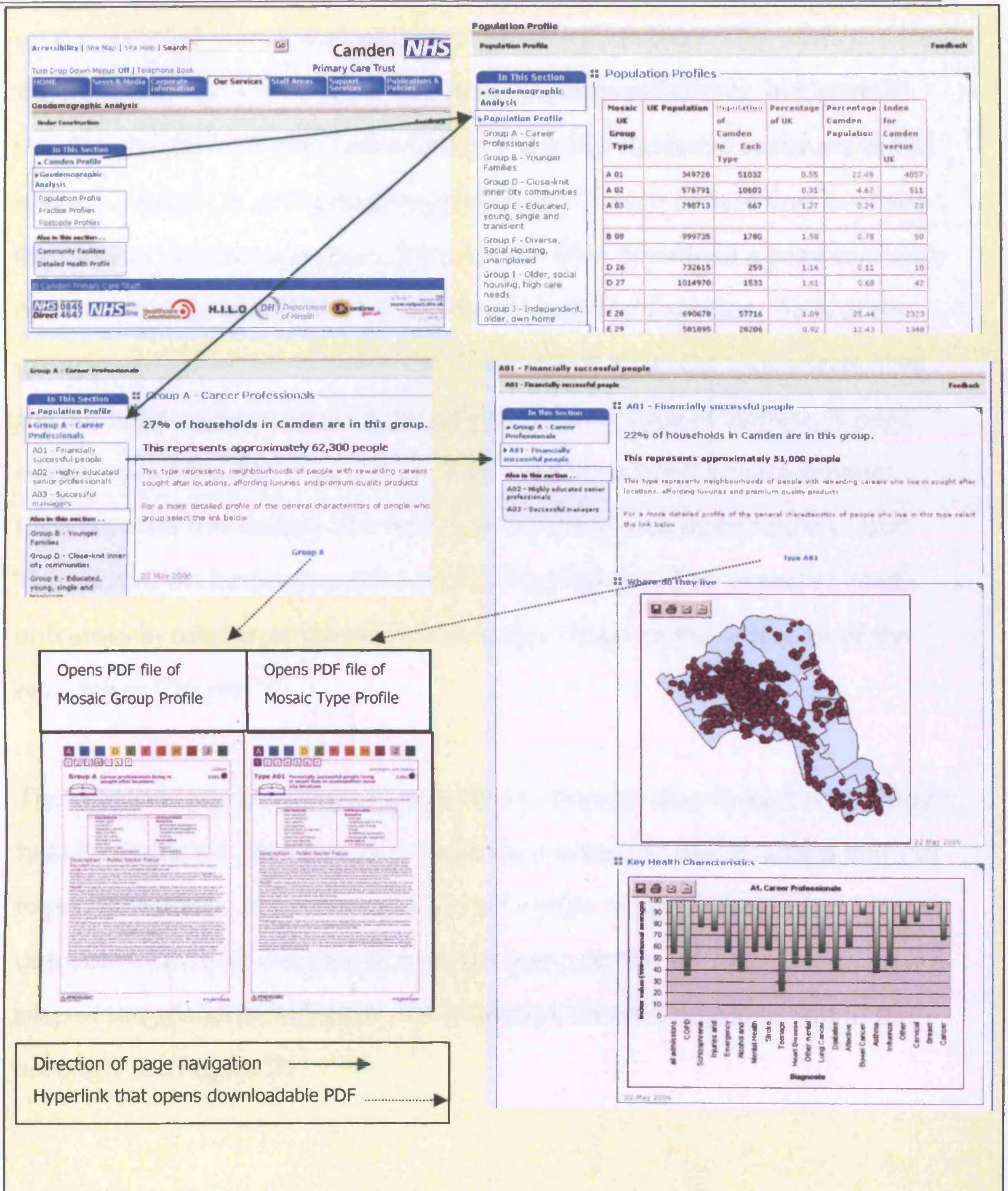


Figure 75: Pages in the population profile of Camden's neighbourhood profile

Figure 75, shows examples of the pages created presenting general neighbourhood characteristics that were extracted from the Mosaic dataset. The user can first explore the general characteristic of the population by comparison across the borough and the UK, by selecting the hyperlink population profiles. The user can then find out about the characteristics of a particular neighbourhood Group, by selecting a Group name from the menu on the left hand side of the population profiles main page. This takes you to the relevant page associated with that neighbourhood Group. In Figure 75, the characteristics of population Group A are highlighted; a summary of the spatial distribution of the neighbourhoods in Camden (where they live) and their key health characteristics. The user can then download a page summary of their more general characteristics as produced by Experian. Each of the groups are further sub-divided into Types. Once again the user selects from the left hand menu to identify the neighbourhood Type of interest. A page opens containing a brief summary of the neighbourhood's characteristics together with two images; the first showing the spatial distribution of that Type across the borough and the second highlighting their expected health outcomes in relation to the national average, based on the outcomes of the research in Chapter 7.

The postcode profiles sub-section provides access to map images for different health conditions. This area of the website enables the user to access maps of expected patterns of health condition for a wide range of diseases and outcomes. The user can choose from the menu on the left hand side to view a map of the spatial distribution of a condition, for each postcode unit in the borough, see Figure 76.

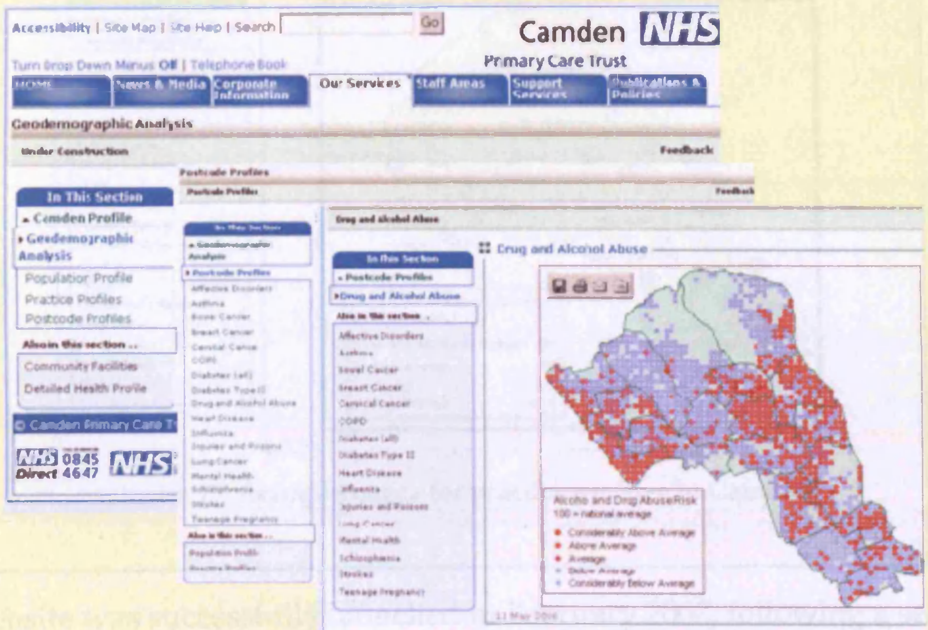


Figure 76: Example pages for postcode profiles in Camden

The third and final section of the website provides access to the practice profiles developed in the case study presented in section 7.2. With the same page structure and navigational ability the practice profiles for each disease or condition were disseminated over the intranet. These maps are HTML image maps so the user can place their mouse curser over a GP practice and the practice code will be displayed.

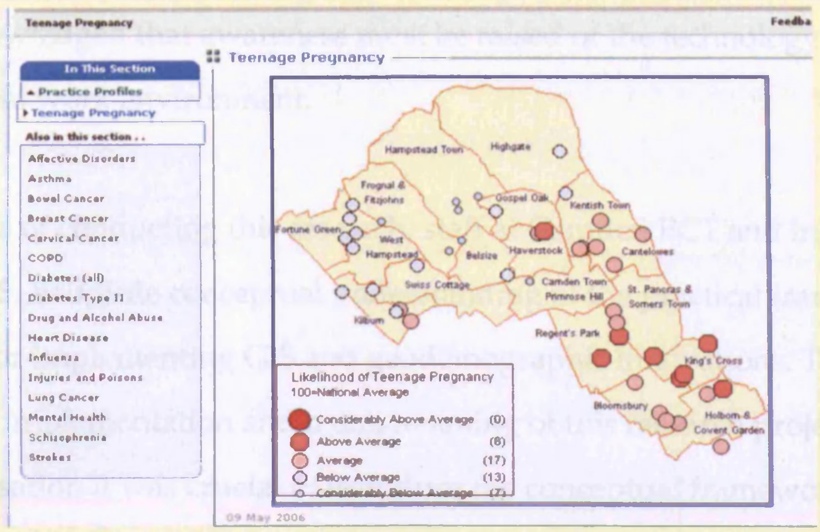


Figure 77: Example pages for practice profiles in Camden

The website was successfully launched in February 2007, following a seminar held in February of the same year to all staff of the Public Health department at Camden. Prior to its official launch a series of knowledge transfer activities (in the forms of case study development, presentations and seminars) occurred throughout the previous year to raise awareness, understanding and embed the geodemographic framework within the PCT. It was hoped that these activities would assist the technology diffusion.

9.3 Communication strategy

The analysis chapters demonstrated practical tools for defining and reaching priority population sub-groups for public health programmes. Whilst the previous section of this chapter discussed the limitations of transferring technology innovations within the NHS and explores workable solutions for this transfer. As noted in section 9.1, technology does not merely refer to machines, and one of its definitions encompasses the notion of scientific knowledge. The development of the Camden Profile was a tangible resource, it was also necessary to undertake knowledge transfer. Alongside the development of the methods, ensuing analysis and dissemination website it was acknowledged that awareness must be raised of the technology within the practical work environment.

At the time of conducting this research, staff at Camden PCT and indeed the wider NHS, had little conceptual understanding of the practical frameworks required for implementing GIS and geodemographic innovations. To ensure successful implementation and mainstreaming of this research project into the organisation it was crucial to introduce the conceptual framework to relevant local public health, public sector and voluntary organisations.

To raise awareness of the conceptual framework of the project I adopted a model used in commercial sector within marketing and brand management, known as the choice pipeline, illustrated in Figure 78. Conceptually this refers to the process of turning a potential customer/user base into a pool of active consumers using an effective chain of marketing and branding steps. In this example, a brand/technology must be promoted to potential consumers/users, which prior to any campaign are unaware of the brand/product. Following marketing, promotional and knowledge transfer activities, the number of customers/users with knowledge of the brand/technology increases and by corollary the number of active customers/users grows, by corollary the number of potential and unaware customers/users decreases. The procedural steps are known as the marketing choice pipeline. In Figure 78 the public health staff are referred to as users and not customers and the website as the product not a brand.

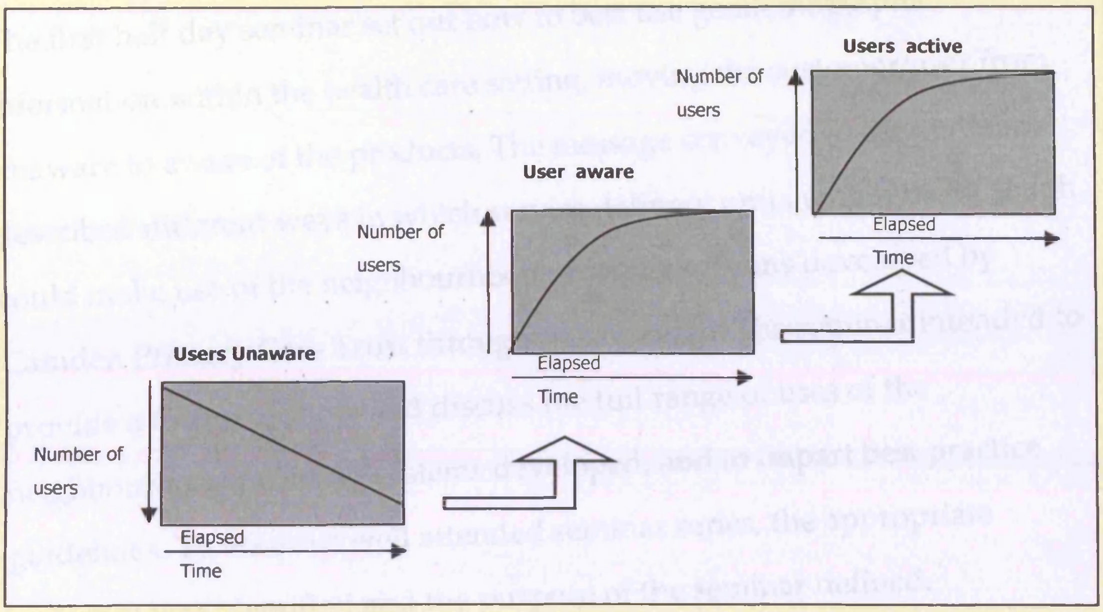


Figure 78: Simplified diagram of the marketing choice pipeline

In this instance, a number of products were being "sold" to a number of users. The products were geodemographics, GIS techniques, and the resultant web-tool. The user base ranged from staff within the public health department, public sector and voluntary workers in the borough, and public health analysts across London.

Embedding new technologies and work channels within any organisation is often difficult and laden with obstacles. The marketing pipeline led to the development of a coherent structure enabling the movement from a user base unaware of the products/technology to an active user base of practitioners understanding and keenly using GIS and Geodemographic techniques within a public health and social marketing setting. A seminar series was the vehicle for which the users were moved through the geodemographics choice pipeline metamorphosing from unaware user to users that were now aware of the products and become active users – building up geodemographic technology knowledge in a professional setting.

9.3.1 Raising awareness through a seminar series

The first half day seminar set out how to best use geodemographic information within the health care setting, moving the customer/user from unaware to aware of the products. The message conveyed to the audience described different ways in which service delivery units within the borough could make use of the neighbourhood profiling systems developed by Camden Primary Care Trust through this research. The seminar intended to provide a forum to show and discuss the full range of uses of the neighbourhood profiling systems developed, and to impart best practice guidelines. To ensure a well attended seminar series, the appropriate audience was identified and the purpose of the seminar defined.

The purpose of the seminar was to introduce various active organisations working within the London Borough of Camden, including the public health department, to the benefits of neighbourhood profiling classifications. The seminar presented the positive outcomes of said profiling system, for improving service delivery and assisting in the numerical quantification of current anecdotal evidence. Outlining its ability to establish local baselines and facilitate monitoring of these baselines. The intended outcome of the seminar was to identify a number of key project champions who would utilise the problem-centred analysis and use it in their day to day activities.

It was not intended to present technical design issues at the seminar, but rather allow professionals working in the field of public service delivery to familiarise themselves with new methods for identifying hard to reach communities. Examples of effective practice within the realm of current practice were given.

The seminar was divided into two parts. The first section centred on the dissemination of geodemographic conceptual theory as applied to identifying priority communities, encompassing a technique known as population profiling. The second section, demonstrated to the audience the practical use of these profiling tools to target communities to improve public health. The list of speakers and the titles of the seminars are listed in Table 28.

Topic	Title	Speaker
Section 1		
Introduction	Overview of Seminar – outline objectives of day	Catherine Jones (UCL)
Priority Communities	Identifying Priority Communities using local & national data	Catherine Jones (UCL)
Priority Communities	The classification of postcodes by surname and forename to provide a proxy indicator of ethnicity.	Pablo Mateos & Richard Webber (UCL)
Section 2		
Targeting Communities	Developing communication strategies for specific services	Suzanne Lutchmun Ana Ileyassoff
Targeting Communities	Slough Diabetes project	Richard Webber
Summary	Drawing together of learning	Richard Webber

Table 28: Agenda of the geodemographics awareness raising seminar

The seminar was advertised using a number of communication channels; word of mouth, poster campaign, personal and colleague network and email mailing list. This ensured that a total of 60 professionals from across public organisations in Camden attended. The results of disseminating this new technology led to a number of staff within the public health department to becoming interested in the geodemographics framework and its application for social marketing and health promotion interventions. Working within the PCT, several off the shelf analysis were carried out within the following fields using the geodemographic framework laid out in section II of the thesis and harnessing the usefulness of the operational health datasets outlined in Chapter 5, section 5.1.2.2:

- In vetro fertilisation (IVF) planning
- Teenage pregnancy
- Stop smoking
- Diabetes awareness
- Flu pandemic planning
- Access to breast screening
- Access to cervical screening

These case studies were conducted for the PCT and were used within their strategy documents to provide visual illustrations of differing health needs of the local population. The actual application to social marketing campaigns was severely limited due to the slow implementation of the social marketing practices recommended by the public health white paper (2004). Every effort was made to present geodemographics as a useful vehicle for social marketing campaigns in Camden. Its successful technology diffusion was severely hindered by the misalignment in the local public health strategy, public policy and the ensuing time lag between the new policy being communicated and its subsequent uptake at the local level.

The research objectives outlined in this thesis were aligned to new government policy associated with the need to use social marketing techniques in bring about public behaviour change (DOH, 2004). The time lag and slow adoption of technology noted by Higgs and Gould (2003, 2005) re-emerged and was evident at Camden PCT. Despite efforts to transfer research knowledge into local and professional knowledge, the lack of both a local social marketing strategy or a spatial data strategy proved considerable stumbling blocks to the technology diffusion this research set out to implement, this can mostly be attributed to the slow prioritisation of social marketing practice at Camden, which is best summarised by these words: "We don't do social marketing in Camden", (Health Promotion Co-ordinator, personal communication: August 2006).

The lack of progress with respect to GIS and spatial needs that was observed in the three years working at the PCT can be attributed to a number of major constraints which hindered progress: lack of a spatial strategy and budget, lack of central database or server or budget to commission one, very fast turnover of staff resulting in loss and constant change in professional knowledge of staff, no current GIS champion (the GIS champion moved on to become the Assistant Director at the Department of Health and no one took over the role left at Camden for two years) within the PCT and a constantly changing working environment resulting from regular issuing of new government policy, targets and restructuring.

9.4 Summary and conclusions

This chapter set out to consider the mechanisms for knowledge transfer, moving the empirical research framework developed in section II into the professional public health domain to build professional knowledge of geodemographics for health professionals and the research field. The social interactionist technology diffusion framework was adopted ensuring that prior to knowledge transfer key considerations were made regarding the situational context in which the diffusion was to occur. Utilisation of the interactionist framework encompassed the recognition of the social context within which the technology was to be implemented, ensuring a user-centred design, sensitivity to change and awareness of cultural uniqueness. It was possible to recognise that the Public Health department operated under considerable technical IT constraints and relevant skills associated with GIS and geodemographics were severely limited. GIS had not moved further on from the simple production of choropleth maps for inclusion in strategic reports. So from this experience collaboration with external organisations can be used effectively to introduce new technologies to organisations, but without the internal champions to actually adopt and use the technologies, it is difficult and challenging to successfully implement new technologies.

With these constraints in mind, a static map repository was developed to disseminate information via the off-the-shelf Intranet site. With the effective communication strategy, the website has now become an important tool for the information team. Although it is not yet used to inform social marketing campaigns it is the "first port of call for information requests. Any requests we have for analysis or information are first directed to the Camden Profile" (Health Intelligence Team Leader, Camden PCT, personal communication: Dec 2007).

The instability and uncertainly that surrounds the working practices at Camden PCT meant the diffusion of new technologies was not simple, and whilst in the number of active number of users is low, there is certainly a greater awareness of the discriminatory power of geodemographic technologies. This can be considered one step nearer to the transference of research knowledge to professional knowledge.

Chapter 10

Conclusions

10 Conclusion

10.1 Reflecting back on the research aims and objectives

This juncture marks the closing chapter of this thesis. It looks back on a journey that started out by justifying the need to develop an alternative framework for measuring and disseminating health outcome inequalities for social marketing through the adoption of a geodemographic approach. It has ended with an alternative working framework for measurement adding value to commonplace operational health data and the widely used national survey data related to health. The research culminated in the dissemination of the framework to an inner London public health department to be implemented as part of a social marketing strategy. Along the way considerations were made for both the accuracy of geodemographics and the application to exploring the complexities of social and geographical scale arising from the geographical analysis conducted.

In the course of the research enough evidence has been produced to validate the decision to produce an alternative framework for health needs and inequalities measurement using geodemographics. Certainly geodemographic classifications, representative of the functional scales of neighbourhoods, when combined with a variety of health datasets provide a rich and detailed picture of the micro-spatial scale in which health inequalities manifest in today's diverse UK society. It has been proven that geodemographics classifications can serve as both flexible and scalable measures of health outcomes for inequality, a valuable new resource for social marketing campaigns and interventions. The thesis presents a collection of research methods and techniques, implemented through a series of case studies that weave together to produce a robust geodemographic framework for informing social marketing campaigns, which was the aim of this thesis.

The result of the research was a set of comparative measures that can be routinely applied to local neighbourhoods in the UK, not just for the study area of Camden presented in this research. This moves forward previous research on the development of such measurements, where efforts have been traditionally hampered by commonalities such as the scale at which data are made available, the limits of small numbers resulting from areal unit aggregations of diseases and outcomes and the issues associated with the modifiable areal unit problem and not withstanding the legal constraints of medical confidentiality and the 1998 Data Protection Act. Although the research was not without its restrictions and the detailed discussion conducted in Chapter 8 raised a number of identifiable limitations of the research but it also noted a few areas worthy of future development, discussed in the subsequent sections of this chapter.

The measures developed during the course of completing section II contributed to achieving the first two objectives of this research thesis and were identified following an extensive literature review. The essence of this thesis was inspired by a multidisciplinary approach; it amalgamated the epistemology and methodological foundations of reading from public health, health inequalities and epidemiology, foundations in sociology, social marketing and marketing, demography, geodemography and geographical analysis. The main body of the thesis, section II, developed and implemented the geodemographic framework for measuring health which was further substantiated by appraising its applicability and replicability to real world issues faced by public health professionals, such as screening initiatives and diabetes awareness. This section also illustrated and substantiated how the geodemographic framework could be used to predict levels of health status and outcomes across different population sub-groups highlighting the most suitable locations for targeting valuable and much needed rationed health resources. The resulting neighbourhood distributions substantiated the view of this thesis that it is pertinent to consider a move away from the lowest decile approach to resource distribution and campaign targeting, as is most commonly utilised when adopting traditional conceptualisations of deprivation and their associated measures. Moving away from the bottom decile approach can be achieved with a drive towards greater awareness and proactive use of preventative public health care initiatives informed by the techniques developed in this thesis. Furthermore, the developed measures indicated neighbourhood levels of health status and their associated health outcomes varied according to neighbourhood types but not specifically by income or deprivation (with the exception of smoking), and so traditional views of poor health associated to poor people is much too simple a conceptualisation and requires refinement in accordance with the specific lifestyle, condition or disease under investigation.

The research also set out to review how the developed geodemographic framework could present the variability of health at the micro-spatial scale whilst accounting for the formal organisational scales that drive all aspects of the operational NHS. Furthermore, the geodemographic framework was extended beyond merely considering health-harming lifestyles to include its application to chronic disease awareness in the form of diabetes and cancer screening, moving into preventing some of the main contributors to diseases of comfort. The derived neighbourhood measures proved themselves to be both flexible and robust enough to be manipulated with a high degree of validity to produce secondary comparable measures of predicted health outcomes for two different organisational scales: the general practice and the larger areal unit of the PCT boundary.

The remaining and final research objective of this thesis was to disseminate the developed framework and research knowledge into the public health setting; creating a potentially new source of professional knowledge by transferring research knowledge into professional knowledge. Chapter 9 addressed this research objective. The success of realising this research objective was somewhat equivocal, but despite the ambiguity some important observations make a positive contribution to the research field. In a constantly changing policy led NHS, there is a considerable time-lag between policy being conceived, communicated, implemented and adopted on the ground by health practioners. Camden PCT was initially slow to take a proactive role in developing the new policy concepts linked to social marketing and its associated campaigns meaning they were slow to adopt the geodemographic framework and its best practice. Despite this, the dissemination of the framework and knowledge carried out using a social interactionist approach was mindful of technology utilisation and its limitations and so its actual diffusion into the department was certainly successful. The knowledge transfer was conducted in such a way that as within the PCT there is re-emergence of the recognition of the potential of GIS and geodemographics, led by another change in management, and these resources are again becoming widely recognised, so much so that staff are now in a position to pick up the professional knowledge transferred and build it up themselves, simply because a suitable foundation was laid.

10.2 Reflection on the underlying research problem

In Chapter 4 where the research framework for the thesis was presented, the multi-faceted nature of the research domains was evident. Readings synthesised existing bodies of research across social marketing and social capital, health inequalities, geodemographics and health measurements. Readings around health inequalities and social marketing acknowledged the requirement for a more comprehensive determinant model to incorporate a population (or neighbourhood) approach to conceptualising patterns of health outcomes. The research developed in this thesis centred on the population sub-group by utilising geodemographic neighbourhood Types to conceptualise the socially similar neighbourhood and in turn calculate the disparate health outcomes most likely to be experienced by the resident population.

Enhancing the relative power of geodemographics by ascribing health characteristics as defined by operational and survey data builds local knowledge relating to the unique characteristics of a place. This means a place becomes more than an artificial construct or abstract form of reality. The framework has contributed by providing some elements of the real context that is essential for understanding everyday communication, social exchange and decision-making, which are all fundamental aspects of social marketing. This will provide social marketers with the new opportunities for exploring, in extraordinary detail, the different micro-scale patterns of social and spatial representation.

The applications of the research to the exploration of differing outcomes associated with diseases of comfort irrefutably indicated the suitability of the neighbourhood as a practical scale for observing health-harming behaviours and the collective attributes of the population comprising that neighbourhood. An intrinsic notion behind social exchange is that it occurs at a local scale. The exchange of social and material resources as supported by social exchange theory, emphasises the interaction between humans occurring at small functional scales of social space. Social exchange was recognized as an important element of successful social marketing, and the application of a national classification of the population sub-groups enhanced by local- level data will in turn provide a much finer scale granularity of the local scale than made previously available. This makes a valuable contribution to social marketers research via the provision of detailed segmentations of patterns related to health variations that are most relevant to key current health issues. Without an evidence base such as the one provided by this research, social marketing campaigns would rely upon more formal scales of evidence provided by the organisational units of the NHS. These formal scales are far less representative of the social processes and social spaces of population sub-groups. For the purposes of social marketing this research has shown the most relevant scale for geographic enquiry is the neighbourhood which approximates homogenous population zones and their social spheres of influence.

10.3 Key achievements

Throughout the course of completing the research aims and objectives, a number of key achievements were realised that have central importance for evidence-based policy, strategy and decision making. The core chapters in section II were dedicated to developing and testing the alternative framework from the perspective of diseases of comfort. One key achievement was the overwhelming ability of geodemographics to differentiate local disparities in health for population sub-groups. Consequently, the geodemographic framework outlined and implemented in this research undoubtedly contributes to the exploration and prediction of local indicators of varying health needs and outcomes with real application and relevance to health practitioners. The use of the framework and measures that were developed play a substantial role in the building of a greater local knowledge of diverse communities, neighbourhoods and their health needs together with existing motivations and behaviours. Kotler and colleagues (Kotler et al., 2002, page 5) noted four objectives that successful social marketers wish their audience to achieve; accept a new behaviour, reject a potential behaviour, modify a current behaviour and abandon an old behaviour. The measures developed herein provided health professionals with relevant evidence with which to develop specific and imaginative campaigns which will by corollary empower their audiences to accomplish these behaviour goals. This approach can be related to the recent paradigm shift in health-care policy, whereby finite resources should be more efficiently managed and targeted at people with the greatest need, with services no longer targeted uniformly towards all the population due to corresponding inefficiencies in cost and resources and the premise of rationed health care services.

Moreover, the innovative use of both operational and survey data in this research contributes to the fields of health measurement, health inequalities research and social marketing. The combining of a national geodemographic classification with either national counts of hospital episodes data or counts acquired from national health surveys proved to be resourceful and effective in producing nationally and locally comparable measures of health outcomes and/or status. A notable innovation in the use of data was attained by using the annual Health Survey for England (HSE), which prior to this study had never before been combined with a national postcode unit classification; its use facilitated the exploration of the discriminatory power of geodemographics to identify predictors of local neighbourhood need related to lifestyle and diseases of comfort. Prior to this study, local indictors of health-harming behaviours such as obesity, drinking and smoking were not available at such fine scale granularity.

Often the HSE is written off because it could not provide specific local detail and the general caution of surrounding is representativeness of populations arising from its sampling frame methodology. Consequently, it was not seen by public health professionals as a useful or valid data source. The geodemographic framework and methodology presented in the first chapter of section II introduced a structure which can be readily applied to this type of data source. Firstly aggregating the survey counts by geodemographic Type and then transforming the survey counts so they are more reflective of the geodemographic distribution of England. This turned a frequently dismissed data source into a useful set of facts which were converted into information to assist local knowledge gathering relating to health-harming lifestyles.

One of the most original contributions and achievements of this research was conceived during the final discussion chapter of section II. Amidst the development of the geodemographic framework and its measures, it came to light that geodemographics may have the potential to conceptualise both the geographic and social space, both of which are fundamental to research in the health domain, and are a critical component of both the social determinant models of health inequalities. This was building on work previously conducted by Moon (1990) and Gatrell (1997) in the field of exploring social space and health geographies and new recognition of the potentialities of geodemographics in the field of sociology by Burrows and Gane (2006). The use of the neighbourhood as a sociological construct to represent functional scales of social spheres of influence gives geodemographics the edge over more conventional analysis of augmenting one or more Census variable with a deprivation indicator to explore social space. The limitations of which are confounded by data issues associated with utilising formal NHS organisational units. Geodemographics provide a framework which can emulate differences in social representations, highlighting areas of social similarity and conversely dissimilarity. They assist the researcher by minimising the likelihood of interactions occurring between different neighbourhoods, an important conceptualisation when targeting new initiatives as it will maximise the likelihood of identifying the appropriate population sub-groups.

Geodemographics provided local representations of social similarity reflecting the distribution of neighbourhood Types. When social processes that are closely associated with social spaces and health status were combined with the unique geographical locations it was possible to visualise both the geographical and social space in parallel, this has not been achieved prior to this research. This provided a unique contribution to the research field, detailing an enriched picture of the local neighbourhood and the uniqueness of local places. The geodemographic neighbourhood provided the starting point for delving deeper into the exploration of the local social context.

Using the geodemographic variables provided in the classification it was possible to define sets of usable social resources (capitals) based on Bourdieu's (1984) work; this created a domain called social status. Health status was then defined using lifestyle, diseases and health outcome measures developed from the Health Survey for England and Hospital Episode data. A representation of Camden's social and geographical space for the health domain has great potential, beyond conventional representations of health inequalities, because the visually capture the emerging inequalities and highlight considerable differences in nearest neighbours by emphasising by social and health divides. In one respect this refutes the geographical notion that near objects/events are more related than those at a distance and highlights the complexity that arise from studying the details of local populations. This contributes to the field of social geography by indicating that health and social status can change dramatically over very small spatial distances and these can now be visually represented.

10.4 Implications for policy and practice

The uniqueness of London as a study area in the UK should not be understated, and the focus on Camden provided a looking glass into the complexity of exploring local health needs and outcomes within such a diverse metropolitan laboratory. The objective to develop professional knowledge relating to this research field principally set its policy direction from the beginning. To ensure the research fitted in with current policy objectives and to maintain its active contribution to future debates, it was essential develop a research strategy inline with the existing policy environment of the public health arena.

In 2004, the seminal Wanless review detailed the types of cost effectiveness action that could improve the health of the whole UK population and ultimately reduce health inequalities. The report went on to state the importance of individuals taking ultimate responsibility for their own health, through the facilitation of appropriate support structures provided by health professionals and government alike.

The geodemographic framework was developed using a number of applications completed as a series of case studies. The case studies selected had particular relevance to diseases of comfort and were chosen because of the underlying significance to current the public health debate and policy guided by the Wanless report. Diseases of comfort are mostly attributable to health-harming behaviours and lifestyles of individuals that are reinforced by their social networks and support. The framework developed in this thesis has the potential to directly contribute to providing an evidence-base around the types of different support structures that should be provided for at different scales of responsibility. It can advocate best practice for segmenting populations because, despite the recurring issues of ecological fallacy befitting all analysis of this type, as was seen in the case studies of Chapter 7, it was proven that the measures maintain their comparability whilst changing scale. The framework produced a series of measures that are adaptable to the scales of policy diffusion. As the scale of policy diffusion changes from local, regional to macro so does the type of health professional implementing such policy change from the local social marketers, health strategists right up to high level senior decision-makers. This research presented a scalable framework for the spatial diffusion of social marketing which has the potential inform the implementation of social health campaigns and interventions that are most relevant to social environments. In turn the framework provides policy makers with a new replicable view of health outcomes that could be used to stimulate community interaction with positive health behavioural changes.

At this point, there is one further noteworthy policy implication of the research which came to fruition in the discussion Chapter 8. A reoccurring observation developed from the data-driven health outcomes measures, developed in the course of section II, was the enriched detail they provided above and beyond traditional methodologies that utilised deprivation measures. If policy were to move away from the deprivation led approach, decisions made using a more comprehensive evidence specific to different types of diseases and their associated interventions, alternative mechanisms (policies) could be developed to deliver health services more specific and relevant for specific populations. The framework presented an enriched picture of local neighbourhood health needs extending beyond standard deprivation measures.

There are practical implications for public health with regards to knowledge transfer, in so far as the research was successful at introducing, improving and increasing the understanding of geodemographic classifications and their analytical potentialities. The intranet resource, albeit technically simplistic, was implemented successfully and provides a simple yet interactive mechanism for viewing geodemographic distributions of health related disparities across neighbourhoods in Camden. It provides a portal for accessing knowledge for public health professional when the time is right and social marketing rises up the local priority agenda. The intranet website could be enhanced by extending the interface to consider web 2.0 geographies, utilising a Google Maps or Microsoft Live Local interface, such as the London Profiler application (www.londonprofiler.org), a Google maps tool interface for viewing public sector data. These new tools provide an interface where ease of use is tantamount to functionality and disseminating the research results using a platform such as this may naturally improve usability and usage with significant implications for knowledge transfer.

10.5 Advantages and limitations

The research conducted was comprised of a series of considered steps with the primary aim to develop and explore an alternative framework for measuring and disseminating health outcome indicators of inequality, for use in social marketing and public health interventions. A number of valid conclusions and contributions to the respective research fields have been identified and outlined in the previous sections. Furthermore, a number of advantages of this research can be identified. The usefulness of geodemographic classifications for social marketing is greatly enhanced when they are enriched with relevant health data, as they then provide a specific detailed local view. This goes above and beyond the application of aggregate census data or deprivation measures to explore areas of inequality. This application is further improved by the annual updating of commercial geodemographic classifications with data to incorporate changes in population demographics. Ascribing the classifications with annually collected health data can provide longitudinal views of local health and changing populations, although this is primarily limited by data availability and validity. A further advantage of the methodology described in this research is the ability to create local predictors of health without the need for manipulation of individual datasets. In a world where data confidentiality is of great public concern, this safeguards medical confidentiality and individual anonymity at the same time as producing measures that are useful and disease specific.

Despite the advantages and many achievements, the research is still subject to a number of constraints and limitations which can be classified into a number of different types; the inherent limitations of geodemographics both methodological and practical, ethical dimensions of using such classifications and limitations associated with data quality and availability, of which are discussed next.

Some of the inherent limitations of geodemographic classifications were discussed in the literature review, and by and large the limitations have been presented throughout this research and were topics for discussion in Chapter 8. Firstly, the uncertainties introduced during the technical construction of each classification cannot be quantified. It is difficult to asses the robustness of the input data and the resultant classification. This is particularly true for commercial classifications, whereby it is not yet possible to adequately test them for reliability and validity. This is because these types of tests are based on classical inference (Lee et al., 2005) and would require in-depth knowledge relating to the input variables and their associated weightings. Because the input data are such an unknown assumptions are made about quality of the data which are taken to be valid, but it is pragmatic to recognise the constraints. Ideally one would have liked to have access to the raw data to asses because any truism of the age old adage "rubbish in: rubbish out" (Longley et al, 2001 page 104). The release of open methodology for the publicly available output area geodemographic classification for 2001 Census data would enable its validity and reliability to be tested and is an area for future work.

The difficulties assessing the accuracy and validity of geodemographics could be overcome if there was a move away from stand alone geodemographics and a move towards bespoke classifications. Instead of a "one-size fits all" classification, bespoke classifications tailored to individual application domains would ensure data inputs are more suitable, the weightings of input variables could be specific to the problem be solved and data such as the HSE could be used to contribute to the classification and not simply be ascribed to the classification post development.

Secondly external uncertainties are caused by statistical aggregations; each classification commercial or open source produces statistical averages of clusters, but it is difficult to determine the extent to which they are a fair and true representation of the local population, how typical are the classifications? The limitations associated with producing averages are nothing new, and there are many statistical measures available to indicate the variations in datasets. This also raises a fundamental question relating to accuracy demands for interventions. Without a true measure it of accuracy for the classifications it is difficult to determine the sensitivity of their applications, this has a number of implications. The first suggests that it is difficult to determine the differences in the averages at the scale of postcode or for the larger output area classifications, it would be useful to have confidence measure to attach to the profiles. An area of research certainly worth further exploration would be to consider the extent to which different communities can be recognised, how much do they differ from the average what is the impact of scale in this type of analysis? The second implication of the sensitivity of these classifications corresponds to the accuracy of working practices. It is a complex challenge social marketer's face. Marketing good health is not simple, and social marketers need to present messages without appearing to be big brother and being too intrusive and controlling, and yet at the same time positively engage with hard to reach communities. The demands for accurate information are implicit, because the wrong health message, communicated poorly to the wrong group could have longstanding consequences for social exchange. This raises the question of whether social marketers can work with the same sensitivity thresholds as commercial marketers. The research made some effort to consider the sensitivity of the classification in chapter 8, but it is certainly an area worth exploring in greater detail in the future.

The third type of uncertainty introduced to this type of analysis occurs when geodemographic typologies are applied to other operational datasets, such as the HSE data. Geodemographic attributes are ascribed to the survey responders. In this instance a survey with coverage for England was adjoined to a classification based on national (England, Scotland and Wales) characteristics. Uncertainties are introduced because one cannot corroborate the sensitivity of the national classification with the population for England. This type of uncertainty can be further extrapolated to the local level, to question whether the uniqueness of demographic population sub-groups in London can be summarised using classification produced to represent national characteristics One possible solution to this type of uncertainty would be to create a unique geodemographic classification using the survey data as the input data, the dependency on this would link back to validity of the sampling frame for the survey, the representativeness of its respondents and the accessibility to the raw data.

Alongside these inbuilt uncertainties that researchers and practitioners encounter when using geodemographics are some more traditional limitations that are inherent to all current geodemographic classifications. Currently these types of datasets only provide a picture at one point in time. As they are only a snapshot of the time at the point in which the data were collected, data producers of classifications such as Experian (Mosaic) and CACI (Acorn) attempt to ameliorate this by enhancing common census data with other publicly and privately collected data. The types of datasets they use are public datasets including county court judgements, company shareholders information or private market research surveys which contain survey questions commissioned by large commercial chains (e.g. Tesco). This added data means datasets are more current than if only built on Census data (such as the Output Area classification); they are updated and commercially released annually or bi-annually. Whilst this means the current state of affairs may be known about a local area, there is still no historical analysis, they cannot be used to truly determine how a neighbourhood has changed.

In the case of social marketing this limitation is not inhibitive to their practice. Social marketers are concerned with changing behaviours of the population in the present day, in the similar ways as commercial marketers are, this is because it is likely they will be less involved with the historical elements of social change. They would be concerned with the historical life-course of their local populations, i.e. where populations grew-up as these factors are also relevant to social determinant models of health inequalities (Marmot and Wilkinson, 2004; Dahlgren and Whitehead, 1991; Bartley, 2004 and Dorling et al, 2007). If such factors could be incorporated into geodemographic classification it would provide a very useful addition to understanding local populations and is another are that would be useful for future development.

The other inherent limitation of geodemographics has been discussed in great detail in both the literature review and Chapter 8. Using any geodemographic classifications the researcher has to be aware of the emergence of the ecological fallacy, which occurs in two ways. Firstly, the classification itself is an ecological construct, but because it is based on social similarity and proximity one can surmise that it is more effective than classifications that are constructed for arbitrary administrative units. Secondly, the ecological fallacy manifests itself when inferences about individuals are made from the group averages of the neighbourhood. This is problematic in the traditional sense, and is a longstanding recurring inaccuracy observed by data analysts. Although, it can be argued that it is unrealistic for social marketers to set out reach every different type of individual on every street because it would be too time consuming, inefficient and costly. Thus within the realm of developing measures for social marketing it is satisfactory to use neighbourhood averages alongside the acceptance of the limitations presented by the ecological fallacy.

The ethical dimension of geodemographic classifications are not often talked about by geodemographers, but it does present another limitation of this research, but at the same time introduces an interesting challenge for future research development. The descriptors used in neighbourhood Types and Groups were created in isolation to the communities they are summarising. This means as descriptors they can be perceived as being highly subjective and the local communities do not actually see themselves as one type of another – the same can be said for deprivation measures. Do people living in "deprived areas" consider themselves deprived? And what is the effect of labelling people as such? The uses of such classifications presents ethical considerations because the visual representations made publicly available could lead to the stigmatisation of certain communities in society.

The final type of limitation with an ethical dimension arises from the confidentiality constraints associated with using data pertaining to human subjects, but is not restricted to this research and is experienced by all types of data analysis. Issues of small datasets are common and can adversely impact the metrics that are developed or expose an individual's identity. In this research it was possible to minimise the effects of small numbers by grouping survey data for a number of years, and as seen in the instance of the breast screening case study data was analysed using the neighbourhood Group and not Type. Geodemographic classifications facilitate the gathering of information relating to social space without compromising an individuals right to data protection or medical confidentiality.

10.6 Thoughts and future directions

A number of limitations were identified throughout the research; some linked to this research in particular and other wider constraints applicable to the wider geodemographic research field. These limitations have great potential for further research development and by identifying and acknowledging these limitations it is possible to envisage which future developments can contribute positively to the fields of geodemographics, social marketing and public health.

On the whole the research conducted using the case studies approach throughout section III presented an exploratory approach to investigating the local population distributions of predicted differences in health outcomes and by corollary their differing health status. One area of work with significant potential to contribute to social marketing and public health might arise from developing the exploratory analysis even further. Measures of spatial autocorrelation essentially measure the tendency of similar values (assigned to locations) to cluster in the spatial domain (Goodchild, 1986) and using such measures would be a useful extension to the geographical analysis presented in section III. Local indictors of spatial autocorrelation such as the LISA statistic (Anselin, 1995) would be useful to measure the extent of local clustering of health outcomes of similar neighbourhoods. Neighbourhoods that exhibit strong positive spatial autocorrelation would inform social marketers of where interventions could naturally be clustered together, making the process even more efficient. It would be useful to expand the exploratory framework into and explanatory model to investigate why health outcomes vary across population sub-groups.

The development of the discussion of scale in health geographies was initiated in this thesis and the application of geodemographic indicators to different scales was considered. It would be appropriate to broaden the discussion and analysis of scale in health geographies to consider local patterns and differences exhibited in the local structure within the context of the influences and interactions between these local structures and the global picture. Furthermore, there is potential for more exploration of the relationships and interactions within the local structure of neighbourhood social similarity and differences of locational proximity.

Although from the research carried out here it seems that prior to any detailed exploratory framework, it would be pertinent to review the limitations of geodemographic classifications that have been identified and develop this area of research first. The following questions extend the current thinking and the limitations relating to geodemographics from a methodological, ethical and application view point, and present them as possible future research questions.

- Can the life-course of populations be incorporated into such classifications?

The life-course of people can be a contributor to social inequalities in health (Marmot and Wilkinson, 2004). Knowing factors such as childhood exposure to good nutrition, medical services, immunisations, education and pollution are important for understanding how health outcomes may be determined at later stages in life. Thus, research into data availability or how appropriate types of data are /could be collected and their subsequent inclusion in geodemographic classifications would enhance their application in the health sector.

- How can geodemographic classifications be developed to account for time-series data used to show historical and social change in local populations?
- Is it possible to have a dynamic geodemographic classification?

Websites such as the intranet site developed in the course of this research have the potential to make available time series data alongside the neighbourhood characteristics, hospital episode data and general practice data, this would go some way to provide a time-series analysis in the short term. Although, neighbourhoods are not static, they are undergoing constant change. Moreover, different neighbourhoods are experiencing regeneration or decline at the same time. If geodemographic classifications could show the changing pattern of social change over time, they could contribute significantly to the social policy and social geography research agendas.

- How can current geodemographic classifications incorporate health data as input variables rather than simply ascribing the health characteristics to existing Types?

It is fair to say that geodemographics provide a more sophisticated approach to population discrimination than IMD and other crude health indictors. The work presented in this thesis suggests that geodemographics work better across the entire social spectrum, this is likely to be because they include a much wider range of more direct indicators of income/wealth. With respect to variable composition, the nature of variable standardisation and the weights assigned to the component variables most commercial geodemographic classifications can be described as black box. More discussion is required on whether the melange of variables that are deemed suitable for discriminating consumption of private goods are actually appropriate for modelling attitudes to service provision, uptake or health behaviour. To supplement the geodemographic classification the coding of the health surveys and analysis conducted in this research is both a practical and useful approach to leveraging the richness of geodemographic profiling for the health sector. In the future research should begin to address the development of bespoke classifications in order to improve the prediction of future health outcomes.

The current one-size fit all applications view to geodemographic classifications is evident, but tailor made classifications for health would provide social marketers with a more enriched picture of the health of their local population and has the potential to greatly enhance the efficiency of campaigns, because the segmentation would be even more specific. With respect to bespoke classifications there are a number of prospects. First, CACI have developed Health ACORN as a sector specific geodemographic classifier but it is still somewhat of a black box; or second, there is potential to rework the OAC classification or other existing classifications in order to accommodate endogenous health data.

Over the recent decades, there has been substantial background development in the power of computer processing, the volumes of both geographic and socio-economic data and the ease at which it is made available (Batty, 2006; Longley et al, 2005). As such the feasibility of assembling relevant datasets and conducting cluster analysis for specific application domains is a realistic option.

- To what extent can the accuracy and validity of such classifications be tested and made publicly available?

The research described in the thesis went some way to exploring how the accuracy and validity of the classifications and is a starting point for further development. It is a relevant research topic because the error margins that are acceptable to commercial marketers may be much wider than those open to social marketers and indeed if classifications are extended to public service resource allocation than it is a fundamental question that needs to, but has not yet been adequately addressed.

- How representative are geodemographics of real communities?

All of these questions relate to the methodological and ethical dimension of geodemographics. In Chapter 8, the Simpson's dissimilarity index was calculated on the HSE variables relating to social class cross-referenced with the neighbourhood Types. The research showed there was actually a lot of dissimilarity in the response counts for different social class within individual neighbourhood Types. This type of analysis could be carried out using more surveys (such as the British Crime Survey, General Household Survey) and available datasets to truly understand the levels of homogeneity with each neighbourhood Type. In the absence of other validity and accuracy measures this has potential to provide a suitable proxy estimate for understanding the actual representativeness of such classifications.

- What do real communities think of how they are defined?
- Can geodemographics classifications be developed ethically to prevent stigmatisation of local communities?

Geodemographic classification can be considered as a data-driven approach to understanding local need and social space. As yet no research has been carried out to understand the implications of describing people in such ways. There has been little research into understanding what real communities think of how they are defined, although Longley and Singleton (2008) have gone some way by gathering user feedback from the E-society classification. More research is necessary to view how real people consider their definitions. It could prove interesting from a sociological perspective to explore how people feel about this data-driven approach to defining them? Furthermore, communities could provide local knowledge which may enhance future classifications. A similar question relates to the ethical implications of these classifications. A report by the Joseph Rowntree Foundation in 2005 (Burrows, 2005) made reference to the new genre of neighbourhood information websites and their potential to widen the gap between rich and poor. Likewise the use of geodemographic classifications may lead to stigmatisation of communities and reinforce social exclusion and marginalisation.

At the beginning of this thesis the research set out to explore and substantiate the development of a geodemographic framework for social marketing. In doing so it was possible to identification and carry out explorative geographical analysis to investigate the spatial distributions of geodemographic measures associated to lifestyle choices and diseases of comfort. As a result the research successfully developed this framework for identifying potential neighbourhoods most likely to be at risk from poor health outcomes and status. In turn this provided a useful and visually powerful means for highlighting neighbourhood variation in likely health outcomes leading to a publicly accessible means of describing the health trajectories of small localities to explore how lifestyle may be represented using geodemographic indicators. The thesis presents a new alternative framework using geodemographics to advance an intelligence led solution for social marketing campaigns.

10.7 Final thought

This thesis has started the process of developing a geodemographic framework that can be applied to real world social marketing application, investigating the linkages between Geodemographic, public health and social marketing research. This is a growing area, with social marketing forging its way into the public policy arena and public health agenda. With public health practitioners slowly recognising the importance of efficiency, effectiveness and tailoring interventions to reach specific community groups, the area can continue to grow stimulated by new methods, frameworks and processes. Furthermore the growth in the area is irrefutable, and is set to continue with companies such as Dr. Foster working in partnership with the NHS Information Centre, to develop best practice for the social marketing process across a range of services and preventative medicine (Farr, Wardlow and Jones, 2008 forthcoming). The need for academic attention to focus of the validity, scalability and accuracy of geodemographic classification together with a framework for practical applications and informed geographical research was demonstrated in throughout this thesis. It is hoped that such research will continue in the near future, due to its important implications for the fields of social marketing, public health and geodemographics.

References

ACTION ON SMOKING AND HEALTH (ASH), 2005. *Smoking and Health Inequalities*, London: ASH. [Online] http://newash.org.uk/files/documents/ ASH_98.pdf [accessed Jan 20th 2008].

ALDER, J., MAYHEW, L., MOODY, S., MORRIS, R., AND SHAH, R., 2005. *The chronic disease burden – An analysis of health risks and health care usage*, CASS Business School: City University. [Online] http://www.cass.city.ac.uk/media/stories/resources/CDB_Oct05.pdf [accessed 19th December 2007].

ANDREASEN, A.R., 2000. *Ethics in Social Marketing*. Washington: Georgetown University Press.

ANDREASEN, A.R., 2003. The life trajectory of social marketing, *Marketing Theory*, 3, 3, pages 293–303.

ANSELIN, L., 1995. Local indicators of spatial association - LISA, *Geographical Analysis*, 27, 2, pages 93 to 116.

ARAYA, R., DUNSTAN, F., PLAYLE, R., THOMAS, H., PALMER, S., AND LEWIS, G., 2005. Perceptions of social capital and the built environment and mental health. *Social Science and Medicine*, 62, 12, pages 3072 to 3083.

ASHBY, D. I., AND LONGLEY, P. A., 2005. Geocomputation, geodemographics and resource allocation for local policing. *Transactions in GIS*, 9, 1, pages 53 to 72.

ASSOCIATION OF GEOGRAPHIC INFORMATION (AGI), 2007, *Glossary of terms*, London: AGI. [Online] http://www.geo.ed.ac.uk/agidexe/term?66 [accessed 4th November 2007].

ASTHANA, S., GIBSON, A., MOON, G., DICKER, J. AND BRIGHAM, P 2004. The demographic and social class basis of inequality in self-reported morbidity: an exploration using Health Survey for England, *Journal of Epidemiology and Community Health*, 58, 4, pages 303 to 307.

ATKINSON, P.M. AND TATE, N.J., 2000. Spatial scale problems and geostatistical solutions: a review, *Professional Geographer*, 52, 4, pages 607 to 623.

BAGOZZI, R.P., 1975. Marketing as Exchange, *Journal of Marketing*, 39, 4, pages 32 to 39.

BAILEY, T.C. AND GATRELL, A.C., 1995. *Interactive spatial data analysis*, Harlow: Longman Scientific & Technical.

BALLAS, D., CLARKE, G., DORLING, D., RIGBY, J., AND WHEELER, B., 2006. Using Geographical information systems and micro-simulation for the analysis of health inequalities, *Health Informatics Journal*, 12, 1, pages 65 to 79.

BANKS, E., BERAL, V., CAMERON, R., HOGG, A., LANGLEY, N., BARNES, I., BULL, D., REEVES, G., ENGLISH, R.,

TAYLOR, S., ELLIMAN, J. AND LOLE HARRIS, C.,2002. Comparison of various characteristics of women who do and do not attend for breast cancer screening. *Breast Cancer Research*, 4, 1. [Online] http://breast-cancer-research.com/content/4/1/R1 [accessed 07/04/2008].

BARTLEY, M., 2004. *Health inequality: an introduction to theories, concepts and methods*, Cornwall: Polity Press.

BARTLEY, M., AND BLANE, D., 1994. Commentary: the appropriateness of deprivation indices must be ensured, *British Medical Journal*, 309, 6967, page 1479.

BATEY, P. AND BROWN, P., 1995. From Human Ecology to Customer Targeting: the Evolution of Geodemographics. In Longley, P.A., Clarke, G.P., (eds), *GIS for Business and Service Planning*, Cambridge: Geoinformation.

BATTY, M., 2006, Public sector information: chains of added value, *Environment and Planning B: Planning and Design*, 33, pages 163 to 164.

BAUDER, H., 2002. Neighbourhood effects and cultural exclusion. *Urban Studies*, 39, 1, pages 85 to 93.

BAUM, F. E., AND ZIERSCH, A. M., 2003. Social capital, *Journal of Epidemiology and Community Health*, 57, pages 320 to 323.

BAUM, F., 1999. Social capital: is it good for your health? Issues for a public health agenda, *Journal of Epidemiology and Community Health*, 54, 6, pages 195 to 196.

BBC, 2006, Smoking 'accounts for health gap'.[Online] http://news.bbc.co.uk/1/hi/health/5176822.stm [last accessed 6 July 2008]

BENZEVAL, M., JUDGE, K., AND WHITEHEAD, M., 1995. *Tackling inequalities in health - an agenda for action*, London: Kings Fund Publishing.

BERKAMN, L., AND GLASS, T., 2000. Social Integration, Social Networks, Social Support and Health. In Berkman,L., and Kawachi, I., (eds), *Social Epidemiology*, Oxford: Oxford University Press.

BERKMAN, L., AND KAWACHI, I., 2000. *Social epidemiology*. Oxford: Oxford University Press.

BIRKIN, M., AND CLARKE, G., 1998. GIS, Geodemographics and Spatial Modelling in the UK Financial Service Industry. *Journal of Housing Research*, 9, 1, pages 8 to 111.

BLANE, D., BRUNNER, E., AND WILKINGSON, R. G., 1996. *Health and social organization: towards a health policy for the twenty-first century*, London: Routledge.

BOLINA, K., LINDRENA, B., LINDSTROMB, M. AND NYSTEDT, P., 2003. Investments in social capital—implications of social interactions for the production of health. *Social Science and Medicine*, 56, 12, pages 2379 to 2390.

BOULOS, M.N.K., 2004. Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom, *International Journal of Health Geographics*, 3, 1. [Online] http://www.ij-healthgeographics.com/content/pdf/1476-072X-3-1.pdf?mkt=2353 [accessed 25th April 2008].

BOURDIEU, P., 1986. *The forms of capital*, New York: Greenwood Press.

BOURDIEU, P., 1997. The forms of capital. In Halsey, A.H., Lauder, H., Brown, P., and Wells, A.S., (eds), *Education: Culture, Economy, Society*, Oxford: Oxford University Press.

BOYLE, P., NORMAN, P., AND REES, P., 2002. Does migration exaggerate the relationship between deprivation and limiting long-term illness? A Scottish analysis, *Social Science and Medicine*, 55, 1, pages 21 to 31.

BROWN, N., 2001. Robert Park and Ernest Burgess, *Urban Ecology Studies*, page 1925.

BROWN, P. J. B., HIRSCHFIELD, A.F.G., AND BATEY, P. W. J., 1991. Applications of geodemographic methods in the analysis of health condition and incidence data. *Journal of the Regional Science Association International*, 70, pages 329 to 344.

BROWN, P. J. B., HIRSCHFIELD, A.F.G., AND BATEY, P. W. J., 2000. *Adding value to census data: Public sector applications of the super profiles geodemographic typology*. [Online] http://www.car.busmgt.ulst.ac.uk/papers/brown.pdf [accessed 13 March 08].

BRUNSDON, C., FOTHERINGHAM, S., AND CHARLTON, M., 1998. Geographically weighted regression-modelling spatial non-stationarity, *The Statistician*, 47, 3, pages 431 to 443.

BURROWS, R., AND GANE, N., 2006. Geodemographics, Software and Class, *Sociology*, 40, 5, pages 793 to 812.

BURROWS, R., ELLISON, N., AND WOODS, B., 2005. *Neighbourhoods on the Net: The nature and impact of internet-based neighbourhood information systems*, Abingdon: Policy Press (Joseph Rowntree Foundation). [Online] http://www.jrf.org.uk/knowledge/findings/housing/0405.asp [accessed 26th December 2007].

BUTTIMER, A., 1969. Social Space in interdisciplinary perspective, *Geographical Review*, 59, 3, pages 417 to 426.

CACI, 2005. *What is Acorn?* London: CACI [Online] *http://www.caci.co.uk/acorn/whatis.asp* [accessed 7th April 2008].

CAMPBELL, H., 1996. Theoretical Perspectives of the Diffusion of GIS Technologies. GIS Diffusion: The Adoption and Use of Geographical Information Systems. In. Masser, I., Campbell, H., and Craglia. M., *Local Government in Europe*, (eds), Bristol: Taylor & Francis Inc., pages 23 to 48.

CAMPBELL, H., AND MASSER, I., 1995. *GIS and Organizations*. London: Taylor & Francis.

CARSTAIRS, V., AND MORRIS, R., 1989. Deprivation, mortality and resource allocation, *Community Medicine*, 11, pages 364-372.

CARSTAIRS, V., AND MORRIS, R., 1991. *Deprivation and health in Scotland*, Aberdeen: Aberdeen University Press.

CATTEL, V., 2001. Poor people, poor places, and poor health: the mediating role of social networks and social capital, *Social Science and Medicine*, 52, 10, pages 1501 to 1516.

CHAINEY, S., AND RATCLIFFE, J., 2005. *GIS and Crime Mapping*, London: John Wiley and Sons Ltd.

CHOI, B.C.K., HUNTER, D.J., TSOU, W., AND SAINSBURY, P., 2005. Diseases of comfort: primary cause of death in the 22nd century, *Journal of Epidemiology Community Health*, 59, pages 1030-1034.

CHUNG, K., YANG, D., AND BELL, R., 2004. Health and GIS: Toward Spatial Statistical Analysis, *Journal of Medical Systems*, 28, 4, pages 349 to 360.

CLAY WAYMAN, J.J., BEALL, T., THACKERAY, R., AND MCCORMACK BROWN, K, R., 2007. Competition: A Social Marketer's Friend or Foe? *Health Promotion Practice*, pages 134 to 140.

COLEMAN, J. S., 1990. *Foundations of social theory*. Cambridge: Harvard University Press

COLEMAN, J. S., 1998. Social capital in creation of human capital. *The American Journal of Sociology*, 94, pages 95-120.

CONGDON, P., 2007. Modelling multiple hospital outcomes: the impact of small area and primary care practice variation, *International Journal of Health Geographics*, 5, 1, pages 50 to 68.

COX, M., BOYLE, P.J., DAVEY, P.G., FENG, Z. AND MORRIS, A.D., 2007, Locality deprivation and Type 2 diabetes incidence: A local test of relative inequalities, *Social Science and Medicine*, 65, pages 1953 to 1964.

CRIME THEORY, 1998. *The Chicago School I: Park and Burgess "Social Ecology"*, Ohio University: Department of Sociology. [Online] http://www.crimetheory.com/Socl/Chicl.htm [accessed 26th December 2007].

CUMMINS, S.C.J., MCKAY, L. AND MCINTYRE, S., 1997. McDonald's restaurants and neighbourhood deprivation in Scotland and England, *American Journal of Preventative Medicine, 29, 4,* pages 308 to 310.

CURRY, M., 1997. The digital individual and the private realm. *Annals of the Association of American Geographers*, 87, pages 681 to 699.

DAHLGREN, G., AND WHITEHEAD, M., 1991. *Policies and strategies to promote social equity in health*, Stockholm: Institute of Future Studies.

DARZI, A., 2007. *Securing our future health: taking a long-term view*, London: HMSO. [Online] http://www.healthcareforlondon.nhs.uk/pdf/aFrameworkForAction.pdf [accessed 19 December 2007].

DAVID, R., 1992. Data access, charging and copyright and their implications for geographical information systems, International *Journal of Geographical Information Science*, 6, 1, pages 13 to 30.

DE SMITH, M. J., GOODCHILD, M. F., AND LONGLEY, P. A., 2008. *Geospatial analysis*, Leicester: Troubador Publishing Ltd. [Online] http://www.spatialanalysisonline.com/ [accessed 20th Jan 2008].

DEBENHAM, J., CLARKE, G., AND STILLWELL, J., 2003. Extending geodemographic classification: a new regional prototype. *Environment and Planning A*, 35, pages 1025 to 1050.

DEDMAN, D., HENNELL, T., HOOPER, J., AND TOCQUE, K., 2006. *Using geodemographics to illustrate health inequalities*, Liverpool: North West Public Health Observatory. [Online] http://64.233.179.104/scholar?hl=en&lr=&client=firefoxa&q=cache:at7ZssssDNsJ:www.nwpho.org.uk/inequalities/Geodem_Inequalities_dr4.pdf+social+marketing+author:dedman [accessed 20th Jan 2008].

DEPARTMENT OF COMMUNITIES AND LOCAL GOVERNMENT (DCLG), 2007, *Indices of deprivation 2007*, London: HMSO. [Online] http://www.communities.gov.uk/communities/neighbourhoodrenewal/deprivation/deprivation07/ [accessed 09 March 2007].

DEPARTMENT OF COMMUNITIES AND LOCAL GOVERNMENT, 2004. *Indices of Deprivation 2004*. London: Neighbourhood Renewal Unit. [Online] http://www.neighbourhood.gov.uk/page.asp?id=1057 [accessed 19th December 2007].

DEPARTMENT OF HEALTH (A), 2004. *Choosing health: making healthier choices easier*. London: HMSO. [Online] http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/Publications PolicyAndGuidance/DH_4094550 [accessed 4th November 2007].

DEPARTMENT OF HEALTH (B), 2004. *National standards local action - health and social care standards and planning framework 2005/06-2007/08*, London: HMSO. [Online] http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/Browsable/DH_4096173 [accessed 30th October 2007].

DEPARTMENT OF HEALTH [C], 2004. *Practice based commissioning: Engaging practices in commissioning*, London: HMSO. [Online] http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicy AndGuidance/DH_4090357 [accessed 20th Jan 2007].

DEPARTMENT OF HEALTH, 1998. *Independent Inquiry into Inequalities in Health Report*. London: HMSO. [Online] http://www.archive.official-documents.co.uk/document/doh/ih/contents.htm [accessed 4th November 2007].

DEPARTMENT OF HEALTH, 2007. *Trends in life expectancy indicator*, London: HMSO. [Online] http://www.communityhealthprofiles.info/indicators/32TrendinLEIndicator.php, [accessed 1st November 2007].

DIEZ ROUX, A.V., 2003. A Glossary for Multilevel Analysis, *Epidemiological Bulletin, 24, 3,* pages 11 to 13

DOOGAN, M. AND ROKKAN, S., 1969. *Quantitative Ecological Analysis in the Social Sciences,* Cambridge: The MIT Press.

DORLING, D., RIGBY J., WHEELER, B., BALLAS, D., THOMAS, B., FAHMY, E., GORDON, D., AND LUPTON, R., 2007. Poverty, wealth and place in Britain, 1968 to 2005, Bristol: The Policy Press.

DRAMOWICZ, K., 2004. A Primer on how to create a customised segmentation system. *Directions Magazine.* [Online] http://www.directionsmag.com/printer.php?article_id=705 [accessed 7th April 2008].

DUNCAN, C., JONES, K. AND MOON, G., 1999. Smoking and deprivation: are there neighbourhood effects? *Social Science and Medicine,* 48, pages 497 to 505.

DURKHEIM, E., 1964. *The division of labour,* New York: Free Press Glencoe.

ENCYCLOPAEDIA BRITANNICA, 2007. *Definition of Species.* [Online] http://www.britannica.com/eb/article-9069027/species [accessed 26th December 2007].

ENCYCLOPAEDIA BRITANNICA, 2007. *Definition of taxonomy.* [Online] http://www.britannica.com/ebc/article-9380336 [accessed 3rd November 2007]

EUROPEAN OBSERVATORY ON HEALTH SYSTEMS AND POLICIES, 2007. *Glossary of Terms,* European Observatory on Health Systems and Policies: WHO European Centre for Health Policy. [Online] http://www.euro.who.int/observatory/Glossary/TopPage?phrase=Health%20outcome [accessed 19th December 2007]

EVERITT, B., 1980. *Cluster Analysis,* New York: Halstead.

EXPERIAN BUSINESS STRATEGIES DIVISION, 2005. *Welcome to the next generation of Mosaic.* [Online] *http://www.business-strategies.co.uk/Content.asp?ArticleID=629* [last accessed 3 November 2006]

FARMER, R., AND LAWRENSON, R., 2004. *Lecture notes on epidemiology and public health medicine.* 5th edition, Oxford: Blackwells.

FARR, M., WARDLAW, J., AND JONES, C.E, 2008. Tackling health inequalities using geodemographics - a social marketing approach, *International Journal of Market Research,* forthcoming.

FIELD, K., 2000. Measuring the need for primary health care: an index of relative deprivation, *Applied Geography,* 20, 4, pages 305 to 332.

FLOWERDEW, R., GRAHAM, E., AND FENG, Z., 2004. *The production of un updated set of data zones to* *incorporate 2001 census geography and data,* Report to the Scottish Executive. [Online] http://www.scotland.gov.uk/Publications/2004/02/18917/33248 [accessed 14 March 08]

FOTHERINGHAM, A.S., BRUNSDON, C. AND CHARLTON, M., 2002. *Quantitative geography: perspectives on spatial data analysis,* London: Sage.

FRENCH, J., AND BLAIR-STEVENS, C., 2006. From snake oil salesmen to trusted policy advisors: The development of a strategic approach to the application of social marketing in England, *Social Marketing Quarterly,* 2, 3, pages 29 to 40.

FRIEDMAN, D. J., HUNTER, E. L., AND PARRISH, R., 2002. *Shaping a health statistics vision for the 21st century,* National Committee on Vital and Health Statistics. [Online] http://www.ncvhs.hhs.gov/hsvision/21st%20final%20report.pdf [accessed 4th November 2007].

GAAG, M., AND SNIJDERS, T., 2004. *The Resource Generator: Social capital quantification with concrete items.* [Online] http://www.xs4all.nl/~gaag/work/RG_paper.pdf [accessed 7th April 2008].

GALOBARDES, B., LYNCH, J., AND SMITH, G. D., 2006. Measuring socioeconomic position in health research, *British Medical Bulletin,* 81, 1, pages 22 to 37.

GATRELL A.C., 2002. *Geographies of health: an introduction,* Oxford: Blackwells Publishing Ltd.

GATRELL A.C., POPAY, J., AND THOMAS, C., 2004. Mapping the determinants of health inequalities in social space: can Bourdieu help us? *Health and Place,* 10, pages 245 to 257.

GEHLKE, C. E., AND BIEHL, K., 1934. Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association,* 29, pages 169 to 170.

GENERAL MEDICAL COUNCIL (GMC), 2004. *Confidentiality: Protecting and Providing Information,* London: GMC. [Online] http://www.gmc-uk.org/guidance/current/library/confidentiality.asp [accessed 1 November 2007].

GLENANE-ANTONIADIS, A., WHITWELL, G., BELL, S.J., AND MENGUC, B., 2003. Extending the vision of social marketing through social capital theory: Marketing in the context of intricate exchange and market failure, *Marketing Theory,* 3, 3, pages 323 to 343.

GOLD, R., KENNEDY, B., CONNELL, F., AND KAWACHI, I., 2002. Teen births, income inequality, and social capital: developing an understanding of the causal pathway, *Health and Place,* 8, pages 77 to 83.

GOLDSTEIN, H., 2003. *Multilevel Statistical Models.* London: Hodder Arnold.

GOODCHILD, M.F., 1986. *Spatial Autocorrelation,* Norwich: Geo Books.

GOODCHILD, M.F., AND PROCTOR, J., 1997. Scale in a digital geographic world. *Geographical and Environmental Modelling,* 1, pages 5 to 23.

GORDON, D., AND PANTAZIS, P., 1997. *Breadline Britain in the 1990's,* Aldershot: Ashgate

GOSS, J., 1995. We know who you are and we know where you live: the instrumental rationality of geodemographic systems. Economic Geography, 71, pages 171 to 198.

GRAHAM, H., 2001. *Understanding Health Inequalities,* Buckingham, Open University Press.

GRUBESIC, T. H., 2004. The geodemographic correlates of broadband access and availability in the United States. *Telematics and Informatics,* 21, pages 335 to 358.

HARPER, R., 2001. *Social capital: A review of the literature, Social Analysis and Reporting Division, Office for National Statistics.* [Online] http://www.statistics.gov.uk/socialcapital/downloads/soccaplitreview.pdf [accessed 1 November 2007].

HARPER, S., AND LYNCH, J., 2006. Measuring health inequalities in Methods. In Oakes, M., and Kaufman, J.S., (eds), *Social Epidemiology,* San Francisco: Jossey-Bass.

HARRIS, R.J, SLEIGHT, P., AND WEBBER, R., 2005. *Geodemographics, GIS and Neighbourhood Targeting,* Chichester: John Wiley and Sons Ltd.

HARRIS, R.J., AND LONGLEY, P.A., 2002. Creating small area measures of urban deprivation, *Environment and Planning A,* 34, pages 1073 to 1093.

HASTINGS, G., 2003. Competition in social marketing, *Social Marketing Quarterly,* 9, 3, pages 6 to 10.

HASTINGS, G., AND SAREN, M., 2003. The critical contribution of social marketing, Theory and application, *Marketing Theory,* 3, 3, pages 305 to 322.

HASTINGS,G., AND MCDERMOT,M., 2006. Putting social marketing into practice, *British Medical Journal,* 332, pages 1210 to 1212.

HAWE, P., AND SHIELL, A., 2000. Social capital and health promotion - a review, *Social Science & Medicine,* 51, 6, pages 871 to 885.

HEISE, D, 2004. Social measurement, classification and scaling. In Smelser, N.J., and Baltes, P. B., (eds), *International Encyclopedia of the Social & Behavioral Sciences,* Oxford: Pergamon Press.

HICKS, J., AND ALLEN, G., 1999. A century of change: trends in UK statistics since 1900. *Social and General Statistics Section: House of Commons Library.* [Online] http://www.parliament.uk/commons/lib/research/rp99/rp99-111.pdf [accessed 19th December 2007].

HIGGS, G., AND GOULD, M.I, 2001. Is there a role for GIS in the 'new NHS'? *Health and Place,* 7, 3, pages 247 to 259.

HIGGS, G., SMITH, D., AND GOULD, M.I, 2003. Realising 'joined-up' geography in the National Health Service: the role of geographical information systems?, *Environment and Planning C: Government and*

Policy, 21, 2, pages 241 to 258.

HIGGS, G., SMITH, D., AND GOULD, M.I., 2005. Findings from a survey on GIS use in the UK National Health Service: organisational challenges and opportunities. *Health Policy*, 72, 1, pages 105 to 117.

HIRSCHFIELD, A. F. G., BROWN, P. J. B., BUNDRED, P., 1995. The Spatial analysis of community health services on Wirral using geographical information systems. Journal of Operational Research, 46, 2, pages, 147 to 159.

HM TREASURY, 2007. *Meeting the aspirations of the British People. 2007 Pre-Budget Report and Comprehensive Spending Review*, London: HMSO. [Online] http://www.hm-treasury.gov.uk/media/7/4/pbr_csr07_completereport_1546.pdf [accessed 27th April 2008].

JARMAN, B., 1983. Identification of Underprivileged areas, *British Medical Journal*, 286, pages 1705-1709.

JARMAN, B., 1984. Underprivileged areas: validation and distribution of scores, *British Medical Journal*, 289, pages 1587 to 1592.

JARVIS, M.J., AND WARDLE, J., 2002. Social patterning of individual health behaviours: the case of cigarette smoking. In Marmot, M., Wilkinson, R., (eds), *Social Determinants of Health*, Oxford: Oxford University Press.

JENERTEE, D.G., AND WU, J., 2000. Defining geographies of scale, *Bulletin of the Ecological Society of America*, 81, 1, pages 104 to 105.

JESSON, J., 2007. Creating a demand for better health by using social marketing techniques. *The Pharmaceutical Journal*, 278, page 776 to 777.

JOCHUM, V., PATTEN, B., WILDING, K., 2005. *Civil Renewal and Active Citizenship: A guide to the debate*, London: National Council for Voluntary Organisations.

JOHNSTON, R. J., GREGORY, D., PRATT, G., AND WATTS, M., 2000. The Dictionary of Human Geography. Oxford: Blackwells Publishing Ltd.

JOHNSTON, R., BURGESS, S., WILSON, D., AND HARRIS, R., 2006. School and residential ethnic segregation: an analysis of variations across England's local education authorities, Regional *Studies*, 40, 9. pages 973 to 990.

JONES, C.E., 2004. *Spatial Distribution of general practice catchments and their markers to deprivation*, MSc Thesis, London: University College London.

JONES, R. A., 1986. Durkheim, Frazer, and Smith: The Role of Analogies and Exemplars in the Development of Durkheim's Sociology of Religion. *American Journal of Sociology*, 92, pages 596 to 62.

KAWACHI, I., SUBRAMANIAN, S.V., ALMEIDA-FILHO, N., 2002. A glossary for health inequalities, Journal of

Epidemiology & Community Health, 56, 9, pages 647-652.

KAWAHCI, I., KENNEDY,B., 1997. The relationship of income inequality to mortality - Does the choice of indicator matter? *Social Science and Medicine*, 45, pages 1121-1127.

KHAN, N, H., AND MUIR, R., 2006. *Sticking together: social capital and local government*, Institute for Public Policy Research and London Borough of Camden. [Online] http://www.camden.gov.uk/ccm/cms-service/stream/asset/?asset_id=590960 [accessed 1 November 2007].

KOTLER, P., LEVY, S.J., 1969. Broadening the concept of marketing, Journal of Marketing, 33, 1, page 10 to 15.

KOTLER, P., ROBERTO, N., AND LEE, N.R., 2002. *Social Marketing: Improving the Quality of Life*, Sage publications, 2nd Edition.

KOTLER, P., ZALTMAN, G., 1971. Social marketing: an approach to planned social change, *Journal of Marketing*, 35, 3, pages 3 to 12.

LAM, N., 2004. Fractals And Scale In Environmental Assessment And Monitoring. In Sheppard, E. and McMaster, R.,(eds), *Scale and Geographic Inquiry: Nature, Society and Method*, Cornwall: Blackwell Publishing.

LANG, L., 2000, Geographies for health organisations, London: ESRI Press.

LAST, J. M., 2001. A dictionary of epidemiology, Oxford: Oxford University Press.

LAYTE, R., WHELAN, C., 2008. Explaining social class inequalities in smoking: the role of education, self-efficacy and deprivation, *European Sociological review*, forthcoming.

LEVENTHAL, B., 1995. Evaluation of geodemographic classifications, *Journal of Targeting, Measurement and Analysis for Marketing*, pages 1-11.

LLOYD, D.A., 2004. *Uncertainty in town centre definition*, doctoral thesis. London: University of London

LOCHNER, K. A., KAWACHI, I., AND KENNEDY, B., 1999. Social capital: a guide to its measurement, *Health and Place*, 5, 4, pages 259 to 270.

LOCHNER, K. A., KAWACHI, I., BRENNAN, R. T., AND BUKA, S. L., 2003. Social capital and neighbourhood mortality rates in Chicago. *Social Science and Medicine*, 56, 8, pages 1797 to 1805.

LONDON SCHOOL OF ECONOMICS, 2006. The Booth Map, London: LSE. [Online] http://booth.lse.ac.uk/static/b/index.html [accessed 1 November 2007]

LONGLEY, P. A., GOODCHILD, M., MAGUIRE, D., RHIND, D., 2001. Geographic Information Systems and Science, New York: John Wiley and Sons.

LONGLEY, P.A., 2005. Geographical Information Systems: a renaissance of geodemographics for public service delivery, *Progress in Human Geography*, 29,1 pages 57 to 63

LONGLEY, P.A., AND BATTY, M., 1996. *Spatial analysis: modelling in a GIS environment*, Cambridge: Geoinformation International

LONGLEY, P.A., AND SINGLETON, A.D., 2008. Classification through consultation: public views of the geography of the e-Society. *International Journal of Geographical Information Science*, 22: in press.

LONGLEY, P.A., GOODCHILD, M., MAGUIRE, D., AND RHIND, D., 2005. Geographical Information Systems and science, John Wiley and Sons, Chichester, 2nd edition

MARMOT, M, AND WIKINSON, R.G., 2004. *Social determinants of health*, Oxford: Oxford University Press.

MARMOT, M., AND BRUNNER, E., 2005. Cohort Profile: The Whitehall II study. International Journal of Epidemiology, 34, pages 251 to 256.

MARMOT, M., AND SHIPLEY, K. J., AND ROSE, G., 1984. Inequalities in death-specific explanations of a general pattern, *Lancet*, 5, pages 1003 to 6.

MARTIN, D., 1996. *Geographic information systems: socioeconomic applications*, New York: Routledge.

MARTIN, D., REES, P., AND WILLIAMSON, P., 2002. *The Census Data System*, Chichester: John Wiley and Sons.

MARTIN, D., TATE, N.J. AND LANGFORD, M., 2000. Refining population surface models: experiments with Northern Ireland Census data, *Transactions in GIS*, 4, 4, pages 434 to 360.

MASSER, I., CAMPBELL, H., AND CRAGLIA. M., 1996., *Local Government in Europe*, Bristol: Taylor & Francis Inc.

MATEOS, P., 2007. *An ontology of ethnicity based upon personal names: with implications for neighbourhood profiling*. PhD Thesis, London: University College London.

MATEOS, P., WEBBER, R., AND LONGLEY , P.A., 2007. *The cultural, ethnic and linguistic classification of populations and neighbourhoods using personal names*, CASA Working Paper 116, University College London: Centre for Advanced Spatial Analysis. [Online] http://www.casa.ucl.ac.uk/working_papers/paper116.pdf [accessed 18th April 2008]

MCCULLOCH, A., 2001. Ward-level deprivation and individual social and economic outcomes in the British Household Panel Study. Environment And Planning A, 33, 4, 667-684.

MCPHERSON, M., SMITH-LOVIN, L., AND COOK, J. M., 2001. Birds of a feather: homophily in social networks. *Annual Review of Sociology*, 27, pages 415 to 44.

MONTELLO, D. R., 2001. Scale in geography. In Smelser, N.J., and Baltes, P. B., (eds), *International Encyclopedia of the Social & Behavioral Sciences*, Oxford: Pergamon Press, pages 13501 to 13504.

MOON, G., 1990. Conceptions of space and community in British health policy, Social Science & Medicine, 30, 1, pages 165 to 171.

MOON, G., QUARENDON,G., BARNARD, S., TWIGG, L., AND BLYTH, B., 2007. Fat nation: Deciphering the distinctive geographies of obesity in England. Social Science & Medicine, 65, pages 20 to 31.

MORGAN, M., MAYS, N., AND HOLLAND, W., 1987. Can hospital use be a measure of need for health care? *Journal of Epidemiology and Community Health*, 41, pages 269 to 274.

NATIONAL CONSUMER COUNCIL, 2005. *National social marketing strategy for health: realising the potential of effective social marketing*, London: HMSO. [Online] HTTP://WWW.NCC.ORG.UK/NCCPDF/POLDOCS/NCC089LT_EFFECTIVE_SOCIAL_MARKETING.PDF [accessed 1 November 2007].

National Social Marketing Centre, 2006. *Social marketing works! A powerful and adaptable approach for achieving and sustaining positive behaviour*, London: HMSO. [Online] HTTP://WWW.NSMS.ORG.UK/IMAGES/COREFILES/NSMC_SOCIAL_MARKETING_WORKS!_LEAFLET_FINAL_NHS_VERSION_DEC2006.PDF [accessed 1 November 2007].

NHS BREAST SCREENING PROGRAMME (NHSBSP), 2005. *NHS breast Screening Programme- annual review.* London: Department of Health. [Online] http://www.cancerscreening.nhs.uk/breastscreen/publications/2005review.html [accessed 18th April 2008].

NHS LONDON, 2006. *Healthcare for London: a framework for action.* London: Department of Health. [Online] http://www.london.nhs.uk/londonnhs-news.aspx?id_Content=7311 [accessed 20th Jan 2008].

NHS, 2007. NHS authorities and trusts. [Online] http://www.nhs.uk/aboutnhs/HowtheNHSworks/authoritiesandtrusts/Pages/authoritiesandtrusts.aspx [accessed 13 March 08].

O'SULLIVAN, D., AND UNWIN, D.J., 2002. *Geographic Information Analysis*, New Jersey: John Wiley and Sons.

OFFICE OF NATIONAL STATISTICS, 2005. *Area classification for Output Areas*, London: HMSO. [Online] http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/oa/default.asp [accessed 26th December 2007].

OFFICE OF NATIONAL STATISTICS, 2005. Methodology, London: HMSO. [Online] http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/oa/downloads/Methodology.pdf [accessed 3rd November 2007].

OFFICE OF NATIONAL STATISTICS, 2008. Life expectancy, London: HMSO. [Online] http://www.statistics.gov.uk/cci/nugget.asp?ID=168 [accessed 6 July 2008].

OFFICE OF PUBLIC SECTOR INFORMATION (OPSI), 1998. Data protection act 1998, London, HMSO. [Online] http://www.opsi.gov.uk/acts/acts1998/ukpga_19980029_en_2#pt1 [accessed 1 November 2007].

OFFICE OF THE DEPUTY PRIME MINISTER (ODPM), 2004. *The English Indices of Deprivation 2004: Summary (revised)*, London: HMSO. [Online] http://www.communities.gov.uk/ archived/publications/communities/indicesdeprivation [accessed 7th April 2008]

OPENSHAW, S., 1984. *The modifiable areal unit problem*, Norwich: Geobooks.

OPENSHAW, S., AND BLAKE, M., 1995. Geodemographic segmentation systems for screening health data. *Journal of Epidemiology & Community Health*, 49, 34 to 38.

OXFORD ENGLISH DICTIONARY, 2002. Concise Oxford Dictionary, Oxford: Oxford University Press

PETO, R., 2000. Smoking, smoking cessation and lung cancer in the UK since 1950: combination of national statistics with two case-control studies, British Medical Journal, 5, 321, pages 323 to 329.

PICKETT, K.E., AND M PEARL, M., 2001. Socioeconomic context and health outcomes: multilevel analysis of neighbourhood, *Journal of Epidemiology & Community Health*, 55, pages 111 to 122.

PORTES, A., 1998. Social capital: Its origins and applications modern sociology. *Annual Review of Sociology*, 24, pages 1 to 24.

PUTNAM, R. D. 1996. The strange disappearance of civic America. *The American Prospect*, 24, winter 1996.

PUTNAM, R. D., 2000. *Bowling alone: The collapse and revival of American community*, New York, Simon and Schuster.

RAPER, J., RHIND, D., AND SHEPHERD, J.W., 1992. *Postcodes: the new geography*, Harlow: Longman Scientific & Technology.

RAPHAEL, D., 2006. Social determinants of health: present status, unanswered questions, and future directions, *International Journal of Health Services*, 36, 4, pages 651 to 77.

REARDON, S., 2006. A conceptual framework for measuring segregation and its association with population outcomes. In Oakes, M.J. and Kaufman, J.S., (eds), *Methods in social epidemiology*, San Francisco: Jossey-Bass for John Wiley and Sons.

REES, P., MARTIN, D., AND WILLIAMSON, P., 2002. The census data system Chichester: John Wiley & Sons.

RITZER, G., 2003. Modern Sociological Theory, London: McGraw-Hill.

ROTHSCHILD, M.L., 1999. Carrots, sticks, and promises: A conceptual framework for the management of public health and social issue behaviours, *Journal of Marketing*, 63, pages 24 to 37

SAMPSON, R. J., RAUDENBUSH, S. W., EARLS, F., 1997. Neighbourhoods and violent crime: a multilevel study of collective efficacy. *Science*, 277, pages 918-924.

SHAW, M., DORLING, D., GORDON, D., AND DAVEY SMITH, G., 1999. *The widening gap: health inequalities and policy in Britain*, Bristol, The Policy Press.

SHEPPARD, E. AND MCMASTER, R., 2003. *Scale and geographic inquiry: nature, society and method*, Cornwall: Blackwell Publishing Ltd.

SINGLETON, 2004. *A state of the art review of geodemographics and their applicability to the higher education market*, London: CASA Working Paper 74. [Online] http://www.casa.ucl.ac.uk/working_papers/paper74.pdf [accessed 26th December 2007].

SLEIGHT, P., 2004. *Targeting customers: how to use geodemographics and lifestyle data in your business*, London: Palgrave Macmillan

SLOGETT, A. AND JOSHI, H., 1994. Higher mortality in deprived areas: community or personal disadvantage? *BMJ*, 309, pages 1470 to 1474.

SMITH, S.J., 2005. Society-Space. In Cloke, P., Crang, P., and Goodwin, M., (eds), *Introducing Human Geographies*,London: Hodder Arnold.

SNOW, J., 1855. On *the mode of communication of Cholera, London*, John Churchill. [Online] http://www.ph.ucla.edu/EPI/snow/snowbook.html [accessed 19th December 2007].

SOCIAL AND SPATIAL INEQUALITIES RESEARCH GROUP, 2005. *The National Classification of Census Output Areas*, Social and Sheffield University: Spatial Inequalities Research Group. [Online] http://www.sasi.group.shef.ac.uk/area_classification/index.html [accessed 26th December 2007].

STAFFORD, M., AND MARMOT, M., 2003. Neighbourhood deprivation and health: does it affect us all equally? *International Journal of Epidemiology*, 32, pages 357-366.

SUBRAMANIAN, S. V., LOCHNER, K. A., KAWACHI, I., 2003. Neighbourhood differences in social capital: a compositional artefact or a contextual construct? *Health and Place*, 9, pages 33 to 44.

TALBOT, R.J., 1991. Underprivileged areas and health care planning: implications of use of Jarman indicators of urban deprivation. *British Medical Journal*, 302, 6773, pages 383 to 386.

THE INFORMATION CENTRE, 2007. *Statistics on smoking: England 2007*, London: HMSO. [Online]http://www.ic.nhs.uk/webfiles/publications/Smoking%20bulletin/Smoking%202007/Statistics%20on%20Smoking%20England%202007.pdf [accessed 19 December 2007].

THE INFORMATION CENTRE, 2007. *The quality and outcomes framework*. London: HMSO [online] http://www.ic.nhs.uk/services/qof [last accessed 20th Jan 2008].

TOBLER, W., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography*, 46, 2, pages 234 to 240.

TONKS, D. AND FARR, M., 1995. Market segments for higher education using geodemographics, *Marketing Intelligence & Planning*, 13, pages 24 to 33.

TONKS, D., 1999. Access to UK higher education, 1991-81: using geodemographics. *Widening participation and lifelong learning*, 17, pages 26 to 36.

TOWNSEND, P., DAVIDSON, N., WHITEHEAD, M., 1988. *Inequalities in health: the black report and the health divide*, London: Penguin

TWIGG, L., MOON, G., 2002. Predicting small area health-related behaviour: a comparison of multilevel synthetic estimation and local survey data, *Social Science & Medicine*, 54, pages 931 to 937.

TWIGG, L., MOON, G., AND JONES, K., 2000. Predicting small-area health related behaviour: a comparison of smoking and drinking indicators, *Social Science & Medicine*, 50, pages 1109 to 1120.

VEENSTRA, G., 2005. Location, location, location: contextual and compositional health effects of social capital in British Columbia, Canada, *Social Science & Medicine*, 60, pages 2059 to 2071.

WAKEFIELD, S.E.L., AND POLAND, B., 2005. Family, friend or foe? Critical reflections on the relevance and role of social capital in health promotion and community development, *Social Science & Medicine*, 60, pages 2819 to 2832.

WALLER, L.A., AND GOTWAY, C.A., 2004, Applied spatial statistics for public health data (Wiley Series in Probability & Statistics), New Jersey: John Wiley & Sons.

WANLESS, D., 2002. Securing our future health: taking a long-term view, London: HMSO. [Online] http://www.hmtreasury.gov.uk/consultations_and_legislation/wanless/consult_wanless_final.cfm [accessed 1 November 2007]

WANLESS, D., 2004. Securing good health for the whole population, London: HMSO. [Online] http://www.hmtreasury.gov.uk/consultations_and_legislation/wanless/consult_wanless04_final.cfm [accessed 1 November 2007]

WEBBER, R., 2004. *Neighbourhood inequalities in the patterns of Hospital Admissions and their application to the targeting of health promotion campaigns.* UCL: CASA Working Paper 90.[Online] http://eprints.ucl.ac.uk/1291/ [accessed 27th April 2008]

WEBBER, R., AND LONGLEY, P.A., 2003. Geodemographic analysis of similarity and proximity: their roles in the understanding of the geography of need. In Longley, P. A., and Batty, M., (eds), *Advanced Spatial Analysis: The CASA book of GIS*, Redlands CA: ESRI Press.

WEBER, R., 2004. *Demographic and Deprivation Ratios: examples of their use in understanding underlying spatial patterns in social phenomena*, UCL: CASA Working Paper 86.[Online] http://www.casa.ucl.ac.uk/publications/workingPaperDetail.asp?ID=86 [accessed 26th December

2007]

WEBER, R., 2004. *The relative power of geodemographics vis a vis person and household level demographic variables as discriminators of consumer behaviour,* UCL: CASA Working Paper 84.[Online] http://www.casa.ucl.ac.uk/publications/workingPaperDetail.asp?ID=84 [accessed 26th December 2007]

WEEKS, J. R. (2004) The Role of Spatial Analysis in Demographic Research. In Goodchild, M.F., and Janelle, D.G., (eds), *CSISS best practice publications: spatially integrated social science,* Oxford: Oxford University Press.

WEGNER, M., AND MASSER, I., 1996. Brave new GIS worlds. In Masser, I., Campbell, H., and Craglia, M., (eds), *GIS Diffusion: The Adoption and Use of Geographical Information Systems in Local Government in Europe.* London: Taylor & Francis.

WIEBE, G.D., 1951. Merchandising Commodities and Citizenship on Television, *The Public Opinion Quarterly,* 15, 4, pages 679to 691.

WOOLCOCK, M., 1998. Social capital and economic development: Toward a theoretical synthesis and policy framework, *Theory and Society,* 27, pages 151 to 208.

World Health Organisation (WHO), 2007. *International Statistical Classification of Diseases and Related Health Problems: 10th Revision version 2007.* [Online] http://www.who.int/classifications/apps/icd/icd10online/ [accessed 20th Jan 2008]

Appendix A

Geodemographic Group/Type[2]	Type of House
Group A – Career professionals living in sought after locations (Symbols of Success)	
People are well set in their careers and their incomes have risen far into upper income tax ranges. Some work for large corporations in senior management positions; some hold respected roles in professional practices; others have built successful enterprises with their own commercial acumen. These are people with busy and complex family lives. Their children are now less time consuming, with more independent lifestyles, but with leisure interests that are likely to be more expensive. This group is mostly white British but is likely to contain significant Jewish, European, Chinese and Indian minorities.	
Type A01 - Financially successful people living in smart flats in cosmopolitan inner city locations	
Type A02 - Highly educated senior professionals, many working in the media, politics and law	
Type A03 - Successful managers living in very large houses in outer suburban locations	
Group B – Younger families living in newer homes (Happy Families)	

[2] All group descriptions taken from : http://www.business-strategies.co.uk/Content.asp?ArticleID=566, last viewed 19th May 2005

Geodemographic Group/Type[2]	Type of House
People whose focus is on career, home and family. They are mostly young couples, married or living with their partner, raising pre-school and school-age children. This group's educational attainment has enabled them to secure positions in large organisations in either the private or the public sector, with the prospect of future career advancement.	
Type B08 - Families and singles living in developments built since 2001	
Group D – Close-knit, inner city and manufacturing town communities (Ties of the Community)	
People live in very established, rather old-fashioned communities. Traditionally, people in this group married young and had manual jobs in industries such as docks and mines. Today, this group has a younger than average population; many are married or cohabiting and bringing up young children. Social support networks are strong, with friends and relations nearby.	
Type D26 - Communities of lowly paid factory workers, many of them of South Asian descent	
Type D27 - Inner city terraced houses attracting second generation Londoners from diverse communities	
Group E – Educated, young, single people living in areas of transient populations (Urban Intelligence)	
Urban Intelligence people are young, well educated and open to new ideas and influences. They are cosmopolitan in their tastes and liberal in their social attitudes. Few have children. Many are in further education while others are moving into full-time employment. Most do not feel ready to make permanent commitments, whether to partners, professions or to specific employers. As higher education has become internationalised, the Urban Intelligence group has acquired many foreign-born residents, which further encourages ethnic and cultural variety.	
Type E28 - Neighbourhoods with transient singles living in multiply occupied large old houses	
Type E29 - Economically successful singles, many living in small inner London flats	

Geodemographic Group/Type[2]	Type of House
Type E30 - Young professionals and their families who have 'gentrified' older terraces in inner London	
Type E32 - Singles and childless couples in small units in newly built private estates outside London	
Type E33 - Older neighbourhoods increasingly taken over by short term student renters	
Type E34 - Halls of residence and other buildings occupied mostly by students	
Group F - People living in social housing with uncertain employment in deprived areas (Welfare Borderline)	
People struggling to achieve the material and personal rewards that are assumed to be open to all in an affluent society. Few hold down rewarding or well-paid jobs; most rely on the council for their accommodation and on state benefits to fund bare essentials. There are high levels of social deprivation in these neighbourhoods. Many tenants are in menial, low-paid jobs and many children live in single-parent families or in transient family formations. In London, a high proportion of this group are of Caribbean or Bangladeshi descent, or have recently arrived in the country as asylum seekers.	
Type F35 - Young people renting hard to let social housing often in disadvantaged inner city locations	
Type F36 - High density social housing, mostly in inner London, with high levels of diversity	
Type F38 - Singles, childless couples and older people living in high rise social housing	
Type F39 - Older people living in crowded apartments in high density social housing	
Type F40 - Older tenements of small private flats often occupied by highly disadvantaged individuals	
Group I – Older people living in social housing with high care needs (Twilight Subsistence)	

Geodemographic Group/Type[2]	Type of House
People have reached the late stage in previously independent lives and now require the support of housing and social services departments. Most rely entirely on state benefits for their income. They either rent their homes from the public sector, rather than owning, or use local authority rather than private care homes. This reflects their low levels of savings and incomes; most do not hold any equity, either in their homes or in financial investments, and their incomes are likely to be restricted to the basic state pension supplemented by other welfare benefit payments.	
Type I48 - People reliant on state pensions living in small council and housing association flats	
Type I49 - Low income older couples renting low rise social housing in industrial regions	
Type I50 - Older people receiving care in homes or sheltered accommodation	
Group J - Grey Perspectives	
People are retired but still independent, with time on their hands and in reasonably good health. They own and run their own homes and are financially independent.	
Type J51 – Elderly people with independent means who have moved to modest apartments	
Type J52 – Well-educated couples and older people in privately owned flats or town houses	

Appendix B

NHS

North Central London Research Consortium

Research Operations Unit

E-mail

Phone
Fax

February 16th 2006

Mr Pablo Mateos & Mrs Catherine Jones
R&D Associates – Public Health Intelligence Team

Dear Mr Mateos & Mrs Catherine Jones,

REC Ref: 05/Q0511/130
Title: Knowledge Transfer Partnership - Geodemographics in Public Health

I am pleased to note that the Local Research Ethics Committee has informed us that there are no ethical reasons why your study should not proceed.

Projects are registered with the North Central London Research Consortium if they utilise patients, staff, records, facilities or other resources of Camden Primary Care Trust, Islington Primary Care Trust, the Camden & Islington Mental Health and Social Care Trust, Barnet Primary Care Trust, Enfield Primary Care Trust or Haringey Teaching Primary Care Trust.

The Camden Primary Care Trust therefore grants approval to begin research based on the proposal reviewed by the ethics committee and subject to any conditions set out in their letter of 29th December 2005. Should you fail to adhere to these conditions or deviate from the protocol reviewed by the ethics committee, then this approval would become void. The approval is also subject to your consent for information to be extracted from your project registration form for inclusion in NHS project registration/management databases and, where appropriate, the National Research Register and the UCL Clinical Research Network register.

Permission to conduct research is also conditional on the research being conducted in accordance with the Department of Health Research Governance Framework for Health and Social Care:

- Appendix A to this letter outlines responsibilities of principal investigators;

- Appendix B details the research governance responsibilities for other researchers. It also outlines the duties of all researchers under the Health and Safety at Work Act 1974. Principal investigators should disseminate the contents of Appendix B to all those in their research teams.

Further information on the research governance framework for health and social care can be found on the DH web pages at http://www.doh.gov.uk/research/
Staff working within trusts covered by the research consortium can also find the information on the Trust Intranet.

Researchers are also reminded that personally identifiable information on living persons must be collected, stored, processed and disclosed in accordance with the Data Protection Act 1998. Such data may be in the form of electronic files, paper files, voice recordings or photographs/scans/X-rays. Further information on the Data Protection Act is available from your organisations Data Protection Officer or from the Consortium R&D Unit. The Medical Research Council also publishes the guidance booklet 'Personal Information in Medical Research' which is available from http://www.mrc.ac.uk/pdf-pimr.pdf

Except in the case of commercially funded research projects, the following acknowledgement and disclaimer MUST appear on all publications arising from your work.
*"This work was undertaken with the support of [***Insert Trust***] Trust, who received [***insert "funding" or a "proportion of funding" ***]* from the NHS Executive; the views expressed in this publication are those of the authors and not necessarily those of the NHS Executive".*

* *"a proportion of funding" where the research is also supported by an external funding body; "funding" where no external funding has been obtained.*

This is a requirement of the contract between the Trust and the NHS Executive in which the Trust receives funding to cover the infrastructure costs associated with performing non-commercial research.

Please make all members of the research team aware of the contents of this approval. I wish you every success with your research.

Yours sincerely,

Research Governance Manager